INAUGURAL-DISSERTATION

zur
Erlangung der Doktorwürde
der
Gesamtfakultät für Mathematik, Ingenieur- und Naturwissenschaften
der
Ruprecht-Karls-Universität
Heidelberg

vorgelegt von

Bianca Marie Neubert
M.Sc. Statistik

aus Bergisch Gladbach

Tag der mündlichen Prüfung:

Nonparametric inference in convolution models

Betreuer: Prof. Dr. Jan Johannes

Acknowledgments

First of all, I would like to express my deepest gratitude to Fabienne Comte for sharing with me her many invaluable insights and knowledge. With seeming endless patience and empathic way of mentoring she steered me through the last years.

Further, I wish to express thanks to the laboratory MAP5 at Université Paris Cité for the financial support and welcoming me for many research stays. I want to thank all of the PhD students who welcomed me open heartedly into their research team.

This endeavor would not have been possible without Jan Johannes and his support, his time and guidance throughout the challenging years of my doctoral studies. I want to thank my colleagues at Heidelberg University who created a pleasant and encouraging work environment. Further, I would like to thank my lectors Sergio, Anna, Maxi and Ricardo for proofreading and helping to reduce the number of typos.

I am sincerely thankful to my family and friends for their encouragement, love, and for always believing in me. Their support has been a constant source of strength and motivation.

I am grateful to the administration of the Combined Faculty of Natural Sciences and Mathematics of the Ruprecht-Karls-Universität Heidelberg. Funding by the Deutsche Forschungsgemeinschaft (German Research Foundation) within the program "Statistical Modeling of Complex Systems and Processes" grant RTG 1953 is also gratefully acknowledged. My thanks go to the Embassy of France in Germany for awarding me the Procope Mobility Grant.

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit der nichtparametrischen statistischen Inferenz in Faltungsmodellen, in welchen die Dichte von Summen oder Produkten unabhängiger Zufallsvariablen, also der additiven bzw. multiplikativen Faltung der einzelnen Dichten, betrachtet wird. Faltungstheoreme liefern, dass sich die zugehörigen Fourier- bzw. Mellin-Transformationen multiplizieren. Diese Eigenschaft kann dazu verwendet werden, statistische Methoden herzuleiten. In dieser Arbeit betrachten wir nichtparametrische Inferenz in zwei verschiedene Faltungsmodellen. Einerseits wird der Fall behandelt, dass die zu schätzende Funktion dem Bild von Dichten unter Faltungen entspricht. Andererseits untersuchen wir das Schätzen quadratischer Funktionale und das Testen von Hypothesen einer unbekannten Dichte bei Vorliegen eines multiplikativen Fehlers. Die Arbeit beginnt mit einer Einleitung und einer Zusammenfassung fundamentaler Methoden im ersten Kapitel und ist anschließend in drei weitere Kapitel geteilt.

Das zweite Kapitel behandelt die Schätzung der p-fachen additiven Faltung der Dichten von p unabhängigen Zufallsvektoren. Zwei nichtparametrische Schätzmethoden werden untersucht: ein Kerndichteschätzer und ein Projektionsschätzer. Das punktweise und integrierte quadratische Risiko wird betrachtet. Mit Hilfe von Fourier-Analysis wird die Varianz beschränkt und klassische Regularitätsklassen erlauben, den Bias zu kontrollieren. Damit werden Konvergenzraten hergeleitet. Die erreichbaren Konvergenzraten hängen von einer optimalen Wahl der Bandweite des Kerndichteschätzers bzw. der Dimension des Projektionsschätzers ab, die in der Praxis nicht bekannt sind. Deshalb stellen wir eine Methode zur Wahl der Bandweite für den Kerndichteschätzer vor und untersuchen Modellwahl für den Projektionsschätzer. Das theoretisch erwartete Verhalten der Schätzer wird in einer Monte-Carlo-Simulationsstudie veranschaulicht.

Das dritte Kapitel der Arbeit thematisiert die Schätzung eines gewichteten quadratischen Funktionals ausgewertet in der Dichte einer strikt positiven Zufallsvariable basierend auf Beobachtungen, die durch einen multiplikativen Fehler verrauscht sind. Wir konstruieren einen Schätzer mit vollständig datengetriebener Wahl des Glättungsparameters basierend auf der Methode von Goldenshluger-Lepski. Anschließend betrachten wir Bedingungen, sodass der Schätzer mit datengetriebener Parameterwahl bis auf logarithmische Faktoren Orakelungleichungen erfüllt. Konvergenzraten werden unter klassischen Regularitätsannahmen hergeleitet. Wir illustrieren die theoretischen Resultate mittels Monte-Carlo-Simulationsstudien.

Das vierte Kapitel adressiert Anpassungstests in multiplikativen Fehlermodellen. Der Abstand, der durch das quadratische Funktional aus Kapitel 3 induziert wird, wird verwendet, um zwischen der Hypothese und der Alternative zu unterscheiden. Wir stellen eine vollständig datengetriebene Teststatistik vor, zeigen obere Schranken für Separationsradien und diskutieren Konvergenzraten unter klassischen Regularitätsannahmen. Die Ergebnisse werden ebenfalls durch Monte-Carlo-Simulationsstudien veranschaulicht.

Die Arbeit schließt mit einem Ausblick.

Abstract

In this thesis, we deal with nonparametric inference for convolution models, which consider the density of the sum or product of real-valued random variables, that is, the additive or multiplicative convolution of the respective densities. Then, convolution theorems yield the multiplication of their corresponding Fourier transforms or Mellin transforms, respectively. These properties are exploited for nonparametric inference. In this thesis, we investigate two types of convolution models. Firstly, we estimate a function which is an image under convolutions. Secondly, we look at the estimation of a quadratic functional and hypothesis testing for an unknown density under multiplicative measurement errors.

After an introduction and a review of fundamental methodologies in the first chapter, the thesis is structured in three further chapters.

The second chapter is concerned with estimating the p-fold additive convolution of the densities of p independent random vectors. Two nonparametric estimators are proposed, a kernel and a projection estimator. Both their pointwise and integrated quadratic risk are investigated. We use Fourier analysis to bound the variance and classical smoothness assumptions to handle the bias, in order to derive rates of convergence. The convergence rates are only attainable if the bandwidth of the kernel estimator or the dimension of the projection estimator, respectively, is optimally selected, which is infeasible in practice. Therefore, we propose a bandwidth selection method for the kernel estimator and study model selection for the projection estimator. The behavior of the estimators is illustrated through Monte-Carlo simulations.

The third chapter deals with the estimation of a weighted quadratic functional evaluated at the density of a strictly positive random variable based on observations corrupted by an independent multiplicative error. We construct a fully-data driven estimator using spectral cut-off regularization. The main theoretical challenge is to establish a data-driven choice of the cut-off parameter using a Goldenshluger-Lepski-method. Conditions under which the fully data-driven estimator attains oracle-inequalities up to logarithmic deteriorations are discussed. We compute convergence rates under classical smoothness assumptions. The results are illustrated using Monte-Carlo simulations.

In the fourth chapter, we cover goodness-of-fit testing in multiplicative measurement error models. The distance induced by the quadratic functional from in the third chapter, is used to distinguish between the null hypothesis and the alternative. We propose a fully-data driven test statistic, show upper bounds for its radius of testing and discuss convergence rates under classical smoothness assumptions. The results are again illustrated through Monte-Carlo simulations.

Perspectives conclude the thesis.

Contents

Introduction

Due to today's technological advancements, we are now able to collect and store large-scale data sets whose complex structures can be modeled in terms of randomness. Consequently, one of the challenges of recent scientific research is to handle such data, which has made statistical learning theory increasingly popular. More precisely, there is a need for the development of elaborate statistical methods which are, on the one hand, dealing with the complexity of data structures and, on the other hand, coping with evolved statistical problems. The aim of this introduction is to motivate the different approaches in the literature for treating these topics and to outline the questions we are addressing in this thesis.

When collecting data, observations are often assumed to be driven by an underlying data generating process unknown to the practitioner and, thus, statistical experiments provide an often widely accepted mathematical formalization. In this thesis, we consider the cases that observations either take values on the real line $\mathbb{R}$, on the positive real line $\mathbb{R}_+$ or are real-valued vectors of some dimension d in the natural numbers $\mathbb{N}$. More precisely, we generally assume to observe a random variable X admitting a probability distribution with density f with respect to the Lebesgue measure λ, in short $X \sim f$. Further, we assume that the density f belongs to some class of functions $\mathscr{F}$.

For example, a typical assumption is that X is normally distributed with unknown mean $\mu \in \mathbb{R}$ and fixed variance equal to one. In this case, each density in $\mathscr{F}$ can be identified with its mean value μ, which is just a real number. If the class of functions $\mathscr{F}$ is restricted in such a way, inference of f, which translates to inference of the unknown real-valued parameter μ, is called parametric. An example of parametric inference is estimating the real-valued parameter μ itself, which is also called point estimation. Then, a mathematical statistician is interested in deriving optimal estimates. There exist multiple concepts of optimality, e.g., minimax and Bayes optimality. A standard reference for point estimation in parametric inference is given by Lehmann and Casella (1998). An alternative objective of statistical inference is hypothesis testing. For an introduction to parametric hypothesis testing, we refer to Lehmann and Romano (2005).

Conversely, inference is called nonparametric if the elements of the class of functions $\mathscr{F}$ cannot be identified with elements of a finite dimensional space. Instead of a parametric form as in the above example, regularity and smoothness assumptions are formulated for the elements of $\mathscr{F}$. In this thesis, classes of interest are Hölder, Nikolski, Sobolev or Mellin-Sobolev spaces which will be introduced at a later point in more detail. For an introduction to nonparametric inference, we refer to Efromovich (1999) or Wasserman (2006), to name but a few.

In nonparametric statistics, estimation and hypothesis testing are possible objectives of inference as well. More precisely, one goal is to derive an estimator $\hat{f}$ of the density f itself based on observations of copies of the random variable X admitting the density f. We refer to the next chapter for a formal definition of an estimator. Then, considering for example densities on $\mathbb{R}$, one is interested in the distance between the estimator $\hat{f}$ and the unknown density f, also called risk or error of estimation. Since, in general, the estimator $\hat{f}$ depends on the observations, which are random, the estimator $\hat{f}$ is in turn random. Thus, a reasonable distance is the expected $\mathbb{L}^1$-distance. Devroye and Györfi (1985) present a detailed definition and discussion of this case, even allowing densities to take values in $\mathbb{R}^d$.

Because of the standard decomposition of the expected quadratic distance into the squared bias and the variance, it provides a more natural distance in statistical inference. At this point let us remark that, since the density f as well as its estimator $\hat{f}$ are functions, there are two ways to evaluate the distance between them: either, at a fixed point, or globally, integrated over the support. One also refers to these cases as point-wise and integrated risks and we refer to Tsybakov (2008)and Comte (2017) for point-wise and global density estimation. For the construction of estimators $\hat{f}$ of the density f, multiple approaches are proposed in the literature. For example, Härdle et al. (1998) discusses inference using wavelets. However, in this thesis, we focus on Kernel and projection approaches.

There exist different notions of optimality in nonparametric inference, out of which minimax optimality is extensively studied in the literature, see Tsybakov (2008). The aim is to derive upper and lower bounds of the risk of an estimator for a given amount $n \in \mathbb{N}$ of independent observations of X, also called sample size. For a fixed sample size n, inference is also called nonasymptotic. An alternative is to consider inference letting the sample size n converge to infinity. We refer to this case as asymptotic statistics and a detailed introduction can be found in van der Vaart and Wellner (1996) or van derVaart (2000).

Until now, we discussed the case of estimation of the density $f \in \mathscr{F}$. Another topic of nonparametric inference vastly examined in the literature is statistical hypothesis testing. Here, we again distinguish between asymptotic and nonasymptotic approaches. For a general, comprehensive introduction to hypothesis testing, we refer to Ingster (1993a,b,c). Ermakov (1994, 1990) covers nonparametric hypothesis testing specifically for densities. A typical null hypothesis is to answer the question whether the distribution of the observed data follows a given fixed density f_0, i.e., $H_0 : f_o = f$, also called goodness-of-fit testing task. To derive test statistics for the goodness-of-fit testing task, a standard method is based on estimating a quadratic distance between the densities f and f_o. Consequently, it is of additional interest to examine the estimation problem of such functionals. For literature on quadratic functional estimation we refer to Bickel and Ritov (1988) and Birgé and Massart (1995). In the case of quadratic functional estimation, we do not make inference for the density f itself but for a quadratic transformation $\theta(f)$ of f. An example for such a transformation is its $\mathbb{L}^2$-norm, i.e., $\theta(f) = \|f\|_{\mathbb{L}^2}$. Then, inference for $\theta(f)$ is still called nonparametric, even if $\theta(f)$ might be real-valued, since its characterization depends on the underlying space $\mathscr{F}$ describing the data-generating density f, which is nonparametric. Other examples of such transformations are linear functionals, e.g., the estimation of the density evaluated at a specific point. There exists extensive literature on linear functional estimation for different models, to name only a few: Lepski and Levit (1998); Butucea (2001); Butucea and Comte (2009); Mabon (2016); Pensky (2017); Brenner Miguel et al. (2023). Moreover, another challenge of inference is when the transformation $\theta(f)$ of the density is again a density or a function belonging to some other function space. An example of such a transformation is the p-fold additive convolution of the density $\theta(f) := f * \cdots * f$. This object appears when considering inference for the distribution of a sum

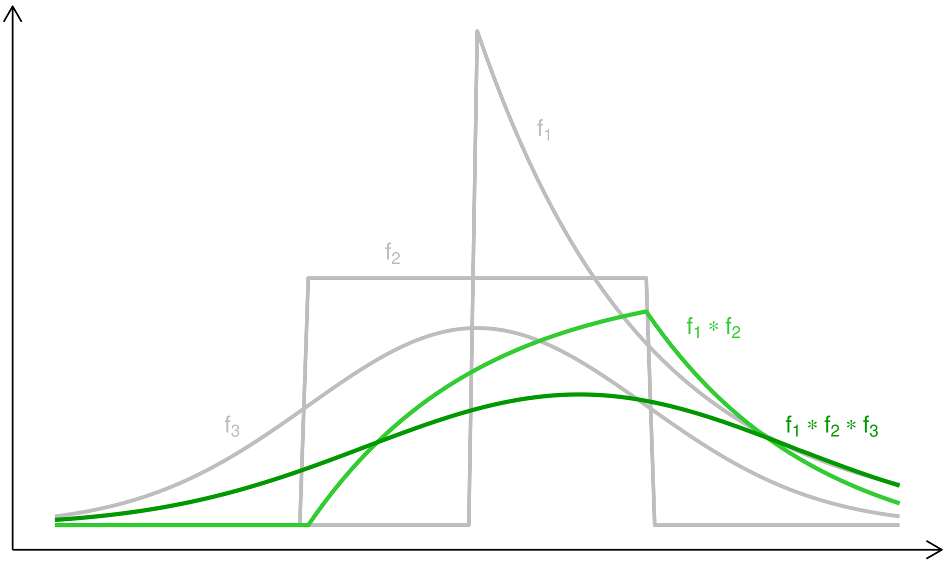


Figure 1: Examples of transformations $\theta(f)$ given by the additive convolution $f_1 * f_2$ (light green) and $f_1 * f_2 * f_3$ (dark green) of marginal densities of $f = (f_1, f_2, f_3)$ (grey lines).

of p independent copies of the random variable X.

Another way to introduce complexity into the statistical model is in form of the observation scheme. If we assume to observe a random variable X admitting an unknown density f and then making inference for some transformation $\theta(f)$, we loosely call this direct inference. However, in practice, we are often not able to observe X directly but only a transformation, e.g., a random variable admitting a density $g = T(f)$ for some transformation T. Making inference for the density f based on Y is also called indirect or (statistical) inverse problem. This is connected to inverse problems, as introduced by Hadamard (1902), see Schluttenhofer (2020) for more details on the connection. Examples for the transformation T are the additive or multiplicative convolution with an error density. Efromovich (1997) and Meister (2009) give an introduction to inference for the additive convolution and Belomestny and Goldenshluger (2020) and Brenner Miguel (2023) for multiplicative convolution. We will discuss such a transformation in more detail in Chapter 1. Let us mention that other transformations are also possible. Tomography, focusing on computerized tomography (CT) and positron emission tomography (PET), is a particular statistical inverse problem that has been studied in recent decades and is related to the Radon transform, see Cavalier (2000, 2001), Korostelev and Tsybakov (1993) and Bissantz et al. (2014).

In the next paragraph we give an outline of the contribution and structure of this thesis.

Contribution and structure of the thesis

The thesis is structured in four chapters: Chapter 1 presents a general review of the underlying mathematical concepts and objects, most importantly the Fourier and Mellin transform, and introduces the statistical models. Chapters 2 to 4 present the statistical results of this thesis.

Roughly speaking, we address two types of complexity of statistical inference. Firstly, in Chapter 2 we consider the estimation of a transformation $\theta(f)$ of f given by the convolution of its marginal densities based on observing independent and identically distributed (i.i.d.) copies of a

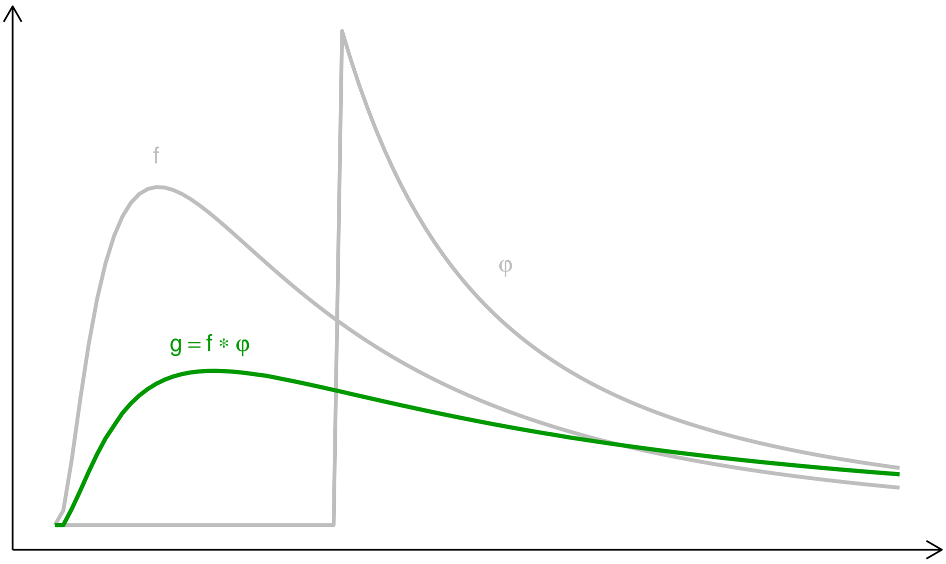


Figure 2: Example of transformation $g = T(f)$ (green line) of the density f by a multiplicative convolution with error density φ (grey lines).

real-valued random vector $X \sim f$. For some density $f = (f_1, f_2, f_3)$, Figure 1 depicts the marginal densities f_1, f_2 and f_3 and the corresponding parameters $\theta(f) = f_1 * f_2$ and $\theta(f) = f_1 * f_2 * f_3$ resulting from the additive convolution of two or three of the marginal densities, respectively. We refer to Chapter 1 for the definition of the additive convolution. In this case, complexity is introduced in form of the transformation θ.

Secondly, in Chapters 3 and 4, complexity is introduced via the transformation T of the density of the data. More precisely, we consider the case of multiplicative convolution with a known error density φ and assume to observe i.i.d. copies of $Y \sim g$ with $g = T(f) = f \circledast \varphi$. The definition of the multiplicative convolution will be given in Chapter 1. An example of densities f and φ and the resulting multiplicative convolution g is depicted in Figure 2. For this model, we consider quadratic functional estimation and the goodness-of-fit testing task.

In general, all chapters start with a summary of proposed results, further references regarding respective models and inference tasks in related literature paragraphs, putting contributions of this thesis into more detailed context. Within sections, main results and necessary definitions are displayed and discussed. Most proofs are presented within the sections, allowing for a deeper understanding of the topic. However, more technical proofs and auxiliary results are postponed to the end of each chapter. Each of the chapters Chapters 2 to 4 end with a simulation study to illustrate the results. Let us summarize the main results of each chapter.

Chapter 2: Estimation for the additive convolution of several multivariate densities Based on i.i.d. observations of independent real-valued random vectors $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(p)}$ admitting densities $f_1, \dots, f_p$, respectively, this chapter is concerned with the problem of estimating the p-fold additive convolution $\theta(f_1, \dots, f_p) = f_1 * \cdots * f_p$ for some $p \in \mathbb{N}$. Two nonparametric estimators are proposed, a kernel and a projection estimator, and their quadratic risk is studied.

For the kernel estimator in Section 2.1, we first show a link between convolution of kernel estimators in the spirit of a plug-in approach and the Fourier formulation. We contribute to the existing literature by extending this estimator to the case of possibly different densities in d dimensions and showing upper bounds of the point-wise and integrated quadratic risk. For both, we discuss suitable regularity assumptions and resulting convergence rates.

In Section 2.2, we contribute to the existing literature the following way: We propose a general projection estimator and prove an upper bound on the integrated quadratic risk. Relying on the Hermite basis, we also build an adaptive estimator using the selection method leading to a parsimonious development in the basis.

– The results of this chapter have been published independently in the preprint Comte and Neubert (2025).

Chapter 3: Quadratic functional estimation in the multiplicative measurement error model This chapter deals with nonparametric estimation of the value of a weighted quadratic functional evaluated at the density f of a strictly positive random variable based on observations $Y \sim g = f \circledast \varphi$ corrupted by an independent multiplicative error φ. Quadratic functionals of the density covered are the $\mathbb{L}^2$-norm of the density, of its derivatives or of the survival function. The contribution to the existing literature can be summarized in the following way. We construct an estimator when the error density is known in Section 3.1 for a fixed cut-off parameter and show upper bounds for its mean squared error in Section 3.2. The main theoretical challenge is the data-driven choice of the cut-off parameter using a Goldenshluger-Lepski-method in Section 3.4. We discuss conditions under which the fully data-driven estimator attains oracle-rates up to logarithmic deteriorations. We compute convergence rates under classical smoothness assumptions.

– The results of this chapter have been published independently in Comte, Johannes and Neubert (2025).

Chapter 4: Adaptive hypothesis testing in the multiplicative measurement error model In this chapter we consider the nonparametric goodness-of-fit testing task for a strictly positive density f based on observations $Y \sim g = f \circledast \varphi$ corrupted by an independent multiplicative error φ. To distinguish between the null hypothesis and the alternative, we consider the distance induced by a weighted norm that is introduced in Chapter 3. We summarize now the contribution to the existing literature. We propose a test statistic in Section 4.1. First, we derive bounds for its quantiles in Section 4.2. Then, we show upper bounds for the radius of testing in Section 4.3. Secondly, we propose a data-driven test statistic using the Bonferroni method and show upper bounds for the radius of testing in Section 4.4. We discuss convergence rates for different collections of cut-off parameters under classical smoothness assumptions.

CHAPTER 1

Methodology

In this chapter, we review for this thesis fundamental methodologies. As we have explained in the introduction, we are interested, on the one hand, in the statistical analysis of a parameter $\theta(f)$ resulting from a p-fold additive convolution of marginal densities, and, on the other hand, in inference based on transformed data with density $T(f)$ resulting from a multiplicative convolution. The analytical tools for both cases are convolution theorems for the additive and multiplicative convolution, respectively. To be able to state these theorems, we introduce the Fourier and Mellin transform. We start by introducing the transforms in Section 1.1 and Section 1.2, respectively. For the Fourier transform we follow the lines of Werner (2018) and Meister (2009). For the Mellin transform we refer to Brenner Miguel (2023).

Afterwards, we review basic notions of nonparametric inference in Section 1.3. This section builds upon results which can be found, for example, in Comte (2017) or Tsybakov (2008) for nonparametric estimation in Section 1.3.1 and Ingster and Suslina (2012) for nonparametric goodness-of-fit testing in Section 1.3.2. In Section 1.4 we introduce the convolution models considered in this thesis.

However, let us first introduce some notation used throughout this thesis.

Notation For some natural number $d \in \mathbb{N}$, let $(\mathbb{R}^d, \mathscr{B}^d, \lambda^d)$ denote the Lebesgue measure space of all d-dimensional real numbers $\mathbb{R}^d$ equipped with the Lebesgue measure λ^d on the Borel σ-field $\mathscr{B}^d$. For $d = 1$, we omit the index d and write $(\mathbb{R}, \mathscr{B}, \lambda)$. Analogously, let $(\mathbb{R}_+, \mathscr{B}_+, \lambda_+)$ denote the Lebesgue measure space of strictly positive real numbers $\mathbb{R}_+$ equipped with the restriction λ_+ of the Lebesgue measure on the Borel σ-field $\mathscr{B}_+$. In contrast, denote by $\mathbb{R}_{\geq 0}$ the positive real line.

Given a density function ν defined on $\mathbb{R}_+$, that is, a Borel-measurable nonnegative function $\nu\colon \mathbb{R}_+ \to \mathbb{R}_{\geq 0}$, let $\nu\lambda_+$ denote the σ-finite measure on $(\mathbb{R}_+, \mathscr{B}_+)$ which is λ_+ absolutely continuous and admits the Radon-Nikodym derivative ν with respect to λ_+. For $p \in [1, \infty]$ let $\mathbb{L}^p_+(\nu) := \mathbb{L}^p(\mathbb{R}_+, \mathscr{B}_+, \nu\lambda_+)$ denote the usual complex Banach-space of all (equivalence classes of) $\mathbb{L}^p$-integrable complex-valued function with respect to the measure $\nu\lambda_+$. Further, for $p \in [1, \infty)$ we define the weighted $\mathbb{L}^p_+(\nu)$-norm of any measurable complex-valued function $h\colon \mathbb{R}_+ \to \mathbb{C}$ by the term

$$\|h\|_{\mathbb{L}_+^p(\nu)}^p := \int_{\mathbb{R}_+} |h(x)|^p \nu(x) d\lambda_+(x) \tag{1.1}$$

and denote by $\|h\|_{\mathbb{L}_+^\infty(\nu)}$ the essential supremum of the function h with respect to $\nu\lambda_+$. For $p = 2$ and $h_1, h_2 \in \mathbb{L}_+^2(\nu)$ denote the corresponding scalar product by

$$\langle h_1, h_2 \rangle_{\mathbb{L}_+^2(\nu)} := \int_{\mathbb{R}_+} h_1(x)\overline{h_2(x)}\nu(x) d\lambda_+(x).$$

If $\nu = 1$, i.e., ν is mapping constantly to one, we write $\mathbb{L}_+^p := \mathbb{L}_+^p(1)$. Analogously, we use the notation $\mathbb{L}_d^p(\nu) := \mathbb{L}^p(\mathbb{R}^d, \mathscr{B}^d, \nu\lambda^d)$ for a density function ν defined on $\mathbb{R}^d$ and define the $\mathbb{L}_d^p(\nu)$-norm and $\mathbb{L}_d^2(\nu)$-scalar product correspondingly. Again, we set $\mathbb{L}_d^p := \mathbb{L}_d^p(1)$. In addition, for $d = 1$, we omit the subscript, i.e., we use the notation $\mathbb{L}^p(\nu) := \mathbb{L}_1^p(\nu)$.

At this point we shall remark, that we have used and will further use the terminology density whenever we are meaning a probability density function (such as f) and on the other hand side density function, whenever we are meaning a Radon-Nikodym derivative (such as ν).

We use the shorthand notation $[\![n]\!] := [1, n] \cap \mathbb{N}$ for any $n \in \mathbb{N}$. Denote by $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ the set of natural numbers including zero. Further, we use multi-index notation relying on bold letters for variables in $\mathbb{R}^d$, that is, $\boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$. For $\boldsymbol{x} = (x_1, \dots, x_d), \boldsymbol{y} = (y_1, \dots, y_d) \in \mathbb{R}^d$ when using standard computations for $\mathbb{R}$, they are to be interpreted component-wise. For example, we write

$$\boldsymbol{x}\boldsymbol{y} = (x_1 y_1, \dots, x_d y_d)$$

and

$$\boldsymbol{x} \leq \boldsymbol{y} \Leftrightarrow (x_k \leq y_k, \forall k \in [\![d]\!]).$$

Taking the sum over $\boldsymbol{j} \in [\![\boldsymbol{m}]\!]$ for $\boldsymbol{m} \in \mathbb{N}^d$ is to be understood as the sum over $j_k \in [\![m_k]\!]$ for all $k \in [\![d]\!]$. Further, we denote the scalar product on $\mathbb{R}^d$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$.

The additive and multiplicative convolution theorem, which will be given below in Property 1.1.2 and Property 1.2.2, respectively, relate the convolution of densities to the product of their Fourier transform and Mellin transform, respectively. Consequently, we give a short review of both transformations, starting with the Fourier transform.

1.1 Fourier transform

The Fourier transform is defined for $h \in \mathbb{L}_d^1$ and extended to $\mathbb{L}_d^2$. Subsequently, we state properties that are essential for the statistical analysis in Chapter 2, following Werner (2018).

The Fourier transform $\mathcal{F}[h]$ of $h \in \mathbb{L}_d^1$ is defined as

$$\mathcal{F}[h](\boldsymbol{t}) := \int_{\mathbb{R}^d} e^{i\langle \boldsymbol{t}, \boldsymbol{x} \rangle} h(\boldsymbol{x}) d\lambda^d(\boldsymbol{x}), \quad \boldsymbol{t} \in \mathbb{R}^d, \tag{1.2}$$

for $i \in \mathbb{C}$ the imaginary unit. Clearly, $\mathcal{F}[h]$ is well-defined. There are alternative definitions with respect to the sign of the argument and the scaling factors $1/\sqrt{2\pi}$ and $1/(2\pi)$, which may be included in the Fourier transform. All results of this thesis can be adapted accordingly leading to modified constants.

The extension of the Fourier transform to $\mathbb{L}_d^2$ is called Fourier-Plancherel-transform and is an isometric isomorphism. Note that for $h \in \mathbb{L}_d^2$ the integral in Equation (1.2) may not exist. However, its Fourier-Plancherel-transformation can be written as an $\|\cdot\|_{\mathbb{L}_d^2}$-limit, integrating over a compact set, we refer to Werner (2018) Theorem V.2.9 for details. If additionally $h \in \mathbb{L}_d^1$, the integral in Equation (1.2) exists and is almost everywhere equal to Fourier-Plancherel-transformation of h. Consequently, in the following, the Fourier-Plancherel-transformation will be identified with the operator $\mathcal{F}$ and we use the terminology Fourier transform for both operators.

First, let us note that Plancherel and Parseval's identity hold for the Fourier transform on $\mathbb{L}_d^2$.

Property 1.1.1:
Let $h_1, h_2 \in \mathbb{L}_d^2$. It holds the *Plancherel equality*

$$\langle h_1, h_2\rangle_{\mathbb{L}_d^2} = \frac{1}{(2\pi)^d}\langle \mathcal{F}[h_1], \mathcal{F}[h_2]\rangle_{\mathbb{L}_d^2}.$$

In particular, the Fourier transform satisfies *Parseval's identity* for all $h \in \mathbb{L}_d^2$, i.e.,

$$\|h\|_{\mathbb{L}_d^2}^2 = \frac{1}{(2\pi)^d}\|\mathcal{F}[h]\|_{\mathbb{L}_d^2}^2.$$

Most importantly, the Fourier transform satisfies a convolution theorem for the additive convolution. For $h_1 \in \mathbb{L}_d^2 \cap \mathbb{L}_d^1$ and $h_2 \in \mathbb{L}_d^2$ the *additive convolution* is defined as

$$(h_1 * h_2)(\boldsymbol{y}) := \int_{\mathbb{R}^d} h_1(\boldsymbol{x})h_2(\boldsymbol{y}-\boldsymbol{x})d\lambda^d(\boldsymbol{x}), \quad \boldsymbol{y} \in \mathbb{R}^d. \tag{1.3}$$

Property 1.1.2 (Additive convolution theorem):
Let $h_1 \in \mathbb{L}_d^2 \cap \mathbb{L}_d^1$ and $h_2 \in \mathbb{L}_d^2$. We have that

$$\mathcal{F}[h_1 * h_2] = \mathcal{F}[h_1] \cdot \mathcal{F}[h_2].$$

In this thesis, we mostly consider the Fourier transform of densities or kernel functions on $\mathbb{R}^d$, which are by definition in $\mathbb{L}_d^1$.

Recall that in Chapter 2 we will consider the estimation of $\theta(f) = f_1 * \dots * f_p$ for the additive p-fold convolution of densities $f_1, \dots, f_p$ on $\mathbb{R}^d$. For some $p \in \mathbb{N}$, the p-fold convolution of functions $h_1, \dots, h_p$ is defined by repetitively taking the convolution, i.e., we have for $\boldsymbol{y} \in \mathbb{R}^d$ that

$$\begin{aligned} &h_1 * \dots * h_p(\boldsymbol{y}) \\ &= \int_{\mathbb{R}^d} \dots \int_{\mathbb{R}^d} h_1(\boldsymbol{x} - \boldsymbol{x}_2 - \dots - \boldsymbol{x}_p)h_2(\boldsymbol{x}_2)\dots h_p(\boldsymbol{x}_p)d\lambda^d(\boldsymbol{x}_2)\dots d\lambda^d(\boldsymbol{x}_p). \end{aligned}$$

The main argument for the construction of the analysis of an estimator in Chapter 2 will be to recursively apply Property 1.1.2. More precisely, we use the resulting equality

$$\mathcal{F}[f_1 * \dots * f_p] = \mathcal{F}[f_1] \cdots \mathcal{F}[f_p].$$

We also have the following properties, which will be used repeatedly in Chapter 2.

Property 1.1.3:
The following properties hold.

(i) Let $h \in \mathbb{L}^2_d \cap \mathbb{L}^1_d$, The Fourier transform $\mathcal{F}[h]$ of h is given by

$$\mathcal{F}[h](\boldsymbol{t}) := \int_{\mathbb{R}^d} e^{i\langle \boldsymbol{t}, \boldsymbol{x}\rangle} h(\boldsymbol{x}) d\lambda^d(\boldsymbol{x}), \quad \boldsymbol{t} \in \mathbb{R}^d,$$

and its inverse $\mathcal{F}^\dagger[h]$ is defined as

$$\mathcal{F}^\dagger[h](\boldsymbol{t}) = \frac{1}{(2\pi)^d}\mathcal{F}[h](-\boldsymbol{t}) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-i\langle \boldsymbol{t}, \boldsymbol{x}\rangle} h(\boldsymbol{x}) d\lambda^d(\boldsymbol{x}), \quad \boldsymbol{t} \in \mathbb{R}^d.$$

(ii) *Young's inequality* states for $1 \le p, q, r \le \infty$ with $\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1$ and $h_1 \in \mathbb{L}^p_d$, $h_2 \in \mathbb{L}^q_d$ that it holds $h_1 * h_2 \in \mathbb{L}^r_d$ and

$$\|h_1 * h_2\|_{\mathbb{L}^r_d} \le \|h_1\|_{\mathbb{L}^p_d}\|h_2\|_{\mathbb{L}^q_d}.$$

(iii) We have boundedness: For $h \in \mathbb{L}^2_d \cap \mathbb{L}^1_d$, it holds that $\|\mathcal{F}[h]\|_{\mathbb{L}^\infty_d} \le \|h\|_{\mathbb{L}^1_d}$.

(iv) If $h \in \mathbb{L}^2_d \cap \mathbb{L}^1_d$ is $\mathbb{R}^d$–valued, then $\mathcal{F}[h](-\boldsymbol{t}) = \overline{\mathcal{F}[h](\boldsymbol{t})}$ for all $\boldsymbol{t} \in \mathbb{R}^d$.

Proofs of Properties 1.1.1 to 1.1.3 can be found in Werner (2018). In particular, we will often be interested in the Fourier transforms of densities on $\mathbb{R}^d$. For a density f on $\mathbb{R}^d$ Property 1.1.3 (iv) can always be applied. In this case, we have by definition that $f \in \mathbb{L}^1_d$ is satisfied. Consequently, assuming additionally $f \in \mathbb{L}^2_d$ is sufficient to have the representations of its Fourier transform $\mathcal{F}[f]$ and inverse given $\mathcal{F}^\dagger[f]$ in (i). Further, for densities, (iii) yields that their Fourier transforms are bounded by one.

While the Fourier transform connects the additive convolution of functions to the product of their Fourier transform, the Mellin transform satisfies an analogous property for the multiplicative convolution. Thus, in the next section, we introduce the Mellin transform and the multiplicative convolution and state important properties.

1.2 Mellin transform

For $c \in \mathbb{R}$ we introduce the density function $\mathrm{x}^c \colon \mathbb{R}_+ \to \mathbb{R}_+$ given by $x \mapsto \mathrm{x}^c(x) := x^c$. The Mellin transform is defined for $h \in \mathbb{L}^1_+(\mathrm{x}^{c-1})$ and extended to $\mathbb{L}^2_+(\mathrm{x}^{2c-1})$ in a similar way the Fourier transform is extended from $\mathbb{L}^1_d$ to $\mathbb{L}^2_d$. We refer to Brenner Miguel (2023) for more details. Here, we define the Mellin transform for square integrable functions following Section 2.3 of Brenner Miguel (2023). Subsequently, we state properties that are essential for the statistical analysis in Chapters 3 and 4.

For $c \in \mathbb{R}$ the Mellin transform $\mathcal{M}_c$ is the unique linear and bounded operator between the spaces $\mathbb{L}^2_+(\mathrm{x}^{2c-1})$ and $\mathbb{L}^2$, which for each $h \in \mathbb{L}^1_+(\mathrm{x}^{c-1}) \cap \mathbb{L}^2_+(\mathrm{x}^{2c-1})$ and $t \in \mathbb{R}$ satisfies

$$\mathcal{M}_c[h](t) = \int_{\mathbb{R}_+} h(x) x^{c-1+2\pi i t} d\lambda_+(x), \tag{1.4}$$

where $i \in \mathbb{C}$ denotes the imaginary unit. Similar to the Fourier transform, there are alternative definitions with respect to the argument and scaling factors and subsequent results can be adapted accordingly. Note that the Mellin transform is defined for functions on the positive real line being integrable with respect to a weighted measure, see Equation (1.1) for the definition.

Analogously to the Fourier transform, the Mellin transform $\mathcal{M}_c$ is unitary, i.e., we have the following property.

Property 1.2.1:
Let $c \in \mathbb{R}$ and $h_1, h_2 \in \mathbb{L}^2_+(\mathrm{x}^{2c-1})$. It holds a *Plancherel-type equality*

$$\langle h_1, h_2 \rangle_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})} = \langle \mathcal{M}_c[h_1], \mathcal{M}_c[h_2] \rangle_{\mathbb{L}^2}.$$

In particular, it satisfies a *Parseval-type identity*, i.e., for all $h \in \mathbb{L}^2_+(\mathrm{x}^{2c-1})$ we have that

$$\|h\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})} = \|\mathcal{M}_c[h]\|^2_{\mathbb{L}^2}.$$

For $h_1 \in \mathbb{L}^1_+(\mathrm{x}^{c-1}) \cap \mathbb{L}^2_+(\mathrm{x}^{2c-1})$ and $h_2 \in \mathbb{L}^2_+(\mathrm{x}^{2c-1})$ the *multiplicative convolution* is defined for $y \in \mathbb{R}_+$ as

$$(h_1 \circledast h_2)(y) := \int_{\mathbb{R}_+} h_1(x) h_2(y/x) x^{-1} d\lambda_+(x). \tag{1.5}$$

Most importantly, the Mellin transform satisfies a convolution theorem for the multiplicative convolution.

Property 1.2.2 (Multiplicative convolution theorem):
Let $c \in \mathbb{R}$, $h_1 \in \mathbb{L}^1_+(\mathrm{x}^{c-1}) \cap \mathbb{L}^2_+(\mathrm{x}^{2c-1})$ and $h_2 \in \mathbb{L}^2_+(\mathrm{x}^{2c-1})$. We have that

$$\mathcal{M}_c[h_1 \circledast h_2] = \mathcal{M}_c[h_1] \cdot \mathcal{M}_c[h_2].$$

Recall that in Chapters 3 and 4, we are interested in inference based on transformed data $g = T(f)$ where T is a multiplicative convolution with a known error density φ. We will use Property 1.2.2 to relate the Mellin transform of the data generating density g with the Mellin transforms of density f and error density φ.

Now, we give a set of Properties and Lemmatas which will be used repeatedly in the statistical analysis of Chapters 3 and 4.

Property 1.2.3:
Let $c \in \mathbb{R}$. The following properties hold.

(i) The Mellin transform's adjoined and inverse $\mathcal{M}_c^\dagger \colon \mathbb{L}^2 \to \mathbb{L}_+^2(\mathrm{x}^{2c-1})$ fulfills for $H \in \mathbb{L}^1 \cap \mathbb{L}^2$ and $x \in \mathbb{R}_+$ that

$$\mathcal{M}_c^\dagger[H](x) = \int_{\mathbb{R}} x^{-c-2\pi i t} H(t) d\lambda(t).$$

(ii) If $h \in \mathbb{L}_+^1(\mathrm{x}^{c-1}) \cap \mathbb{L}_+^2(\mathrm{x}^{2c-1})$ is real-valued, it holds that

$$\overline{\mathcal{M}_c[h]}(t) = \mathcal{M}_c[h](-t), \quad t \in \mathbb{R},$$

and, analogously, for $H \in \mathbb{L}^1 \cap \mathbb{L}^2$ real-valued it holds

$$\overline{\mathcal{M}_c^\dagger[H]}(x) = \mathcal{M}_c^\dagger[H(-\cdot)](x), \quad x \in \mathbb{R}_+.$$

(iii) Let $h_1, h_2 \in \mathbb{L}_+^1(\mathrm{x}^{2c-2})$. Then, $h_1 \circledast h_2 \in \mathbb{L}_+^1(\mathrm{x}^{2c-2})$ with

$$\|h_1 \circledast h_2\|_{\mathbb{L}_+^1(\mathrm{x}^{2c-2})} \leq \|h_1\|_{\mathbb{L}_+^1(\mathrm{x}^{2c-2})} \|h_2\|_{\mathbb{L}_+^1(\mathrm{x}^{2c-2})}.$$

(iv) Let $h_1, h_2 \in \mathbb{L}_+^1(\mathrm{x}^{c-1})$ and $h_2 \in \mathbb{L}_+^2(\mathrm{x}^{2c-1})$. Then, $h_1 \circledast h_2 \in \mathbb{L}_+^1(\mathrm{x}^{c-1}) \cap \mathbb{L}_+^2(\mathrm{x}^{2c-1})$ with

$$\|h_1 \circledast h_2\|_{\mathbb{L}_+^2(\mathrm{x}^{2c-1})} \leq \|h_1\|_{\mathbb{L}_+^1(\mathrm{x}^{c-1})} \|h_2\|_{\mathbb{L}_+^2(\mathrm{x}^{2c-1})}.$$

For proofs of the results stated in Properties 1.2.1 to 1.2.3 we refer to Brenner Miguel (2023). As we will again mostly consider Mellin transforms of densities, let us make a few remarks. First, densities are real valued and (ii) can be applied. Further, if in (iii) h_1, h_2 are additionally densities, equality holds. To be able to use the Mellin transform and the formula of its inverse given in (i), we assume throughout Chapters 3 and 4 that considered densities belong to $\mathbb{L}_+^1(\mathrm{x}^{c-1}) \cap \mathbb{L}_+^2(\mathrm{x}^{2c-1})$ for some $c \in \mathbb{R}$.

Next, we proof an additional norm inequality for the multiplicative convolution.

Lemma 1.2.4:
Let $h_1, h_2 \in \mathbb{L}_+^\infty(\mathrm{x}^{2c-1})$ and $h_2 \in \mathbb{L}_+^1(\mathrm{x}^{2c-2})$ for $c \in \mathbb{R}$. Then, $h_1 \circledast h_2 \in \mathbb{L}_+^\infty(\mathrm{x}^{2c-1})$ and it holds that

$$\|h_1 \circledast h_2\|_{\mathbb{L}_+^\infty(\mathrm{x}^{2c-1})} \leq \|h_1\|_{\mathbb{L}_+^\infty(\mathrm{x}^{2c-1})} \|h_2\|_{\mathbb{L}_+^1(\mathrm{x}^{2c-2})}.$$

Lemma 1.2.4. We see that

$$\begin{aligned}
\|h_1 \circledast h_2\|_{\mathbb{L}_+^\infty(\mathrm{x}^{2c-1})} &= \sup_{y \in \mathbb{R}_+} \left| y^{2c-1} \int_{\mathbb{R}_+} h_1(y/x) h_2(x) x^{-1} d\lambda_+(x) \right| \\
&\leq \int_{\mathbb{R}_+} \sup_{y \in \mathbb{R}_+} |y^{2c-1} h_1(y/x) h_2(x) x^{-1}| d\lambda_+(x) \\
&= \int_{\mathbb{R}_+} |h_2(x)| x^{-1} \sup_{y \in \mathbb{R}_+} |y^{2c-1} h_1(y/x)| d\lambda_+(x).
\end{aligned}$$

By change of variables, we further get

$$\begin{aligned}\|h_1 \circledast h_2\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})} &\le \int_{\mathbb{R}_+} |h_2(x)| x^{-1} x^{2c-1} \sup_{y\in\mathbb{R}_+} |y^{2c-1} h_1(y)| d\lambda_+(x) \\ &= \|h_2\|_{\mathbb{L}^1_+(\mathrm{x}^{2c-2})} \|h_1\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})}.\end{aligned}$$

This concludes the proof. □

Recall that in Chapters 3 and 4, we assume that for some error density φ, we observe some random variable $Y \sim g = f \circledast \varphi$. Denoting by $\mathbb{E}_g$ the expectation with respect to g, we relate the $\mathbb{L}^1_+(\mathrm{x}^{2c-1})$ norm of g to the corresponding moment of Y. More precisely, it holds that

$$\mathbb{E}_g[Y^{2c-1}] = \int_{\mathbb{R}_+} x^{2c-1} g(x) d\lambda_+(x) = \|g\|_{\mathbb{L}^1_+(\mathrm{x}^{2c-1})}.$$

Further, $g = f \circledast \varphi$ is equivalent to $Y = X \cdot U$ in distribution, for $\mathbb{R}_+$-valued random variable $U \sim \varphi$ admitting density φ independent of X. Combining both properties, we relate the $\mathbb{L}^1_+(\mathrm{x}^{2c-1})$-norm of g to the corresponding norms of f and φ, that is, it holds

$$\|g\|_{\mathbb{L}^1_+(\mathrm{x}^{2c-1})} = \mathbb{E}_g[Y^{2c-1}] = \mathbb{E}_f[X^{2c-1}]\mathbb{E}_\varphi[U^{2c-1}] = \|f\|_{\mathbb{L}^1_+(\mathrm{x}^{2c-1})} \cdot \|\varphi\|_{\mathbb{L}^1_+(\mathrm{x}^{2c-1})}. \tag{1.6}$$

We use these computations repeatedly in Chapters 3 and 4. We further deduce the following properties.

Lemma 1.2.5: (i) Let $g \in \mathbb{L}^\infty_+(\mathrm{x}^{2c-1})$, then it holds for $\mathbb{1}_{[-k,k]} h \in \mathbb{L}^2$ that

$$\mathbb{E}_g\left[\left|\int_{-k}^{k} Y^{c-1+2\pi i t} h(t) d\lambda(t)\right|^2\right] \le \|g\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})} \|\mathbb{1}_{[-k,k]} h\|^2_{\mathbb{L}^2}.$$

(ii) For $g \in \mathbb{L}^\infty_+(\mathrm{x}^{2c-1})$ and $h \in \mathbb{L}^2_+(g)$ we have that

$$\int_{\mathbb{R}} \left|\mathbb{E}_g\left[Y^{c-1+2\pi i t} h(Y)\right]\right|^2 d\lambda(t) \le \|g\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})} \mathbb{E}_g[h^2(Y)].$$

(iv) If $f \in \mathbb{L}^1_+(\mathrm{x}^{2c-2})$ and $\varphi \in \mathbb{L}^\infty_+(\mathrm{x}^{2c-1}) \cap \mathbb{L}^1_+(\mathrm{x}^{2c-2})$ are densities. Then, for $g := f \circledast \varphi$ we have that

$$\|g\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})} \le \|f\|_{\mathbb{L}^1_+(\mathrm{x}^{2c-2})} \|\varphi\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})} = \frac{\|\varphi\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})}}{\|\varphi\|_{\mathbb{L}^1_+(\mathrm{x}^{2c-2})}} \|g\|_{\mathbb{L}^1_+(\mathrm{x}^{2c-2})}.$$

Proof of Lemma 1.2.5. For (i) we first see that

$$\begin{aligned}\mathbb{E}_g\left[\left|\int_{-k}^{k} Y^{c-1+2\pi i t} h(t) d\lambda(t)\right|^2\right] &= \int_{\mathbb{R}_+} g(y) y^{2c-1} y^{1-2c} \left|\int_{-k}^{k} y^{c-1+2\pi i t} h(t) d\lambda(t)\right|^2 d\lambda(y) \\ &\le \|g\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})} \int_{\mathbb{R}_+} y^{1-2c} \left|\mathcal{M}^\dagger_{1-c}[\mathbb{1}_{[-k,k]} h](y)\right|^2 d\lambda(y) \\ &= \|g\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})} \|\mathcal{M}^\dagger_{1-c}[\mathbb{1}_{[-k,k]} h]\|^2_{\mathbb{L}^2_+(\mathrm{x}^{1-2c})}.\end{aligned}$$

Finally, with the Parseval-type identity, see Property 1.2.1, it follows

$$\mathbb{E}_g\left[\left|\int_{-k}^{k} Y^{c-1+2\pi i t} h(t) d\lambda(t)\right|^2\right] \leq \|g\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})} \|\mathbb{1}_{[-k,k]} h\|^2_{\mathbb{L}^2}.$$

For (ii), we again apply the Parseval-type identity given in Property 1.2.1 and get that

$$\int_{\mathbb{R}} \left|\mathbb{E}_g\left[Y^{c-1+2\pi i t} h(Y)\right]\right|^2 d\lambda(t) = \|\mathcal{M}_c[gh]\|_{\mathbb{L}^2} = \|gh\|_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})}$$
$$= \int_{\mathbb{R}_+} g^2(y) h^2(y) y^{2c-1} dy \leq \|g\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})} \mathbb{E}_g[h^2(Y)].$$

For (iii), apply Property 1.2.3 (iv), Equation (1.6) and Lemma 1.2.4. This completes the proof of Lemma 1.2.5. □

This concludes the review of the Fourier and Mellin transform. We continue by summarizing basic notions of nonparametric inference in the next section.

1.3 Nonparametric inference

In this section, we review general notions of nonparametric inference which will be considered in this thesis. We start with nonparametric estimation in Section 1.3.1 and continue with hypothesis testing in Section 1.3.2. Let us first introduce some notation.

We assume to observe a random variable X in some measure space $(\mathcal{X}, \mathcal{A}, \mu)$ and write shortly $X \in \mathcal{X}$. As already mentioned in the introduction, in this thesis, we consider the two measure spaces $(\mathcal{X}, \mathcal{A}, \mu) = (\mathbb{R}^d, \mathscr{B}^d, \lambda^d)$ for some $d \in \mathbb{N}$ and $(\mathcal{X}, \mathcal{A}, \mu) = (\mathbb{R}_+, \mathscr{B}_+, \lambda_+)$. We consider the space of probability measures on $(\mathcal{X}, \mathcal{A}, \mu)$ admitting a density with respect to μ and denote by $\mathscr{F}$ the set of these densities. That is, we assume X to admit a density $f \in \mathscr{F}$. We write $X \sim f$ and denote by $\mathbb{P}_f$ its probability distribution and denote by $\mathbb{E}_f$ the corresponding expectation.

Further, for a set $(X_j)_{j\in[\![n]\!]}$ of independent and identically distributed (i.i.d.) copies of X, also called sample of size $n \in \mathbb{N}$, we denote the corresponding product space, measure and expectation by $\mathcal{X}^n$, $\mathbb{P}^n_f$ and $\mathbb{E}^n_f$, respectively.

As already mentioned, we will make inference for different types of transformations $\theta\colon \mathscr{F} \to \Theta$ applied to the density$f$. We consider different image sets Θ depending on the example. For the p-fold convolution of densities on $\mathbb{R}^d$, the image space is again a function space, e.g., $\Theta \subseteq \mathbb{L}^2_d$ or $\Theta \subseteq \mathbb{L}^\infty_d$. When considering quadratic functionals, it holds $\Theta \subseteq \mathbb{R}_+$ and for goodness-of-fit testing, we test for the density itself, i.e., $\Theta \subseteq \mathscr{F}$. The specific transformations will be defined later. Let us first outline the main objectives of inference discussed in this thesis.

1.3.1 Nonparametric estimation

One topic of interest in nonparametric inference is parameter estimation, which we consider in Chapters 2 and 3. This section follows Tsybakov (2008) and Comte (2017).

Any measurable function $\hat{\theta}\colon \mathcal{X}^n \to \Theta$ is called *estimator* of $\theta(f)$. For a sample $(X_j)_{j\in[\![n]\!]}$ of $X \sim f$ we write shortly $\hat{\theta}_n := \hat{\theta}(X_1, \ldots, X_n)$. A standard measure of accuracy of an estimator is its risk. More precisely, for some measurable distance function $\delta\colon \Theta \times \Theta \to \mathbb{R}_{\geq 0}$, we define the risk of estimator $\hat{\theta}$ of $\theta(f)$ for $f \in \mathscr{F}$ as

$$\mathcal{R}_n(\hat{\theta}, \theta(f)) := \mathbb{E}_f^n[\delta(\hat{\theta}_n, \theta(f))].$$

Further, one often is interested in the behavior of the estimator uniformly over the class $\mathscr{F}$. That is, one considers its *maximal risk*, defined as

$$\mathcal{R}_n(\hat{\theta}, \mathscr{F}) := \sup_{f\in\mathscr{F}} \mathbb{E}_f^n[\delta(\hat{\theta}_n, \theta(f))].$$

Depending on the parameter of interest, in this thesis, we consider different distance functions for the risk.

(i) In Chapter 2, the image set of the parameter of interest Θ consists of the densities on $\mathbb{R}^d$ satisfying certain integrability assumptions. We look at two different risks: the integrated quadratic risk and the point-wise quadratic risk. The *integrated quadratic risk* is defined as the risk with loss function

$$\delta(h_1, h_2) = \|h_1 - h_2\|^2_{\mathbb{L}^2_d}, \quad h_1, h_2 \in \mathbb{L}^2_d.$$

The *point-wise quadratic risk* is defined as the risk with the loss function given for $\boldsymbol{x} \in \mathbb{R}^d$ by

$$\delta(h_1, h_2) = |h_1(\boldsymbol{x}) - h_2(\boldsymbol{x})|^2, \quad h_1, h_2 \in \mathbb{L}^\infty_d.$$

(ii) In Chapter 3, the parameter of interest it a quadratic functional, i.e., $\theta\colon \mathscr{F} \to \mathbb{R}_{\geq 0}$, given by $\theta(f) = \|f\|^2$ for some norm specified later. We consider the *mean squared error*, i.e, the risk with respect to the loss function

$$\delta(x, y) = |x - y|^2, \quad x, y \in \mathbb{R}_{\geq 0}.$$

Let $(\psi_n)_{n\in\mathbb{N}}$ be a sequence of values in $\mathbb{R}_+$ such that $\psi_n \to 0$ as $n \to \infty$. We say that an estimator admits a *convergence rate* $(\psi_n)_{n\in\mathbb{N}}$ over regularity class $\mathscr{F}$ if for some $C \geq 0$ it holds

$$\lim_{n\to\infty} \frac{\mathcal{R}_n(\hat{\theta}, \mathscr{F})}{\psi_n} \leq C.$$

In nonparametric estimation, the estimator $\hat{\theta}$ often depends on a tuning parameter $k \in \mathcal{K}$ for some index set $\mathcal{K}$. Consequently, we have a family of estimators

$$\{\hat{\theta}_k, k \in \mathcal{K}\}.$$

Usually, k is interpreted as a smoothing parameter, for example the dimension in projection estimation, the bandwidth in kernel estimation or the spectral cut-off parameter in inverse problems. An optimal choice of the tuning parameter is given by

$$k^*(n, \theta(f)) = \arg\min_{k\in\mathcal{K}} \mathcal{R}_n(\hat{\theta}_k, \theta(f)).$$

Since this parameter usually depends on the regularity of f, which is unknown in practice, the estimator $\hat{\theta}_{k^*}$ resulting from this choice of tuning parameter is called *oracle* for $\mathcal{K}$ with respect to the risk $\mathcal{R}_n$. The following question arises: Can we construct an estimator $\hat{\theta}_*$ that mimics the behavior of the oracle $\hat{\theta}_{k^*}$ in the sense that

$$\mathcal{R}_n(\hat{\theta}_*, \theta(f)) \leq \mathcal{R}_n(\hat{\theta}_{k^*}, \theta(f)) + \frac{1}{n}$$

for a wide range of functions f? In this case, the estimator $\hat{\theta}_*$ is called *data-driven*. We discuss different data-driven approaches in Chapters 2 and 3. Let us now turn to nonparametric hypothesis testing. More precisely, we consider goodness-of-fit testing.

1.3.2 Nonparametric goodness-of-fit testing

In Chapter 4, we consider goodness-of-fit testing. Let us briefly recapitulate some elementary concepts, as introduced in Ingster (1993a,b,c).

Let $\mathscr{F}_0$ be a subset of $\mathscr{F}$ which is non-empty and not equal to $\mathscr{F}$. A *goodness-of-fit testing task* considers the null hypothesis

$$H_0 \colon f \in \mathscr{F}_0.$$

In this thesis, we study a *simple hypothesis*. That is, the set $\mathscr{F}_0 := \{f_o\}$ consists of one element for some $f_o \in \mathscr{F}$. The goal is then to construct a decision rule based on observations $(X_j)_{j\in[\![n]\!]}$ on sample space $\mathcal{X}^n$ following distribution $\mathbb{P}_f$. A decision rule has values zero and one. Here, a value of zero means that the null hypothesis is accepted. A value of one signifies that the null hypothesis is rejected. More precisely, a *decision rule* is defined as a measurable map $\Delta_n \colon \mathcal{X}^n \to \{0, 1\}$ also called *test*. To evaluate the performance of a test, we consider two types of errors. The *type I error probability* is defined as the probability to reject the null hypothesis whenever it is true, i.e.,

$$\mathbb{P}^n_{f_o}(\Delta_n = 1).$$

A natural alternative is given by $H_1 \colon f \in \mathscr{F}_1 = \mathscr{F} \setminus \{f_o\}$. In contrast, the *type II error probability* is defined as the probability of accepting the null hypothesis whenever it does not hold, that is, for $f \in \mathscr{F}_1$ by the probability

$$\mathbb{P}^n_f(\Delta_n = 0).$$

A typical problem is that the alternative set $\mathscr{F} \setminus \{f_o\}$ contains points too close to the point f_o to significantly differentiate between null hypothesis and alternative, roughly speaking. As solution to this problem one removes some neighborhood of the point of f_o from the alternative. More precisely, we look for some distance function $\delta \colon \mathscr{F} \times \mathscr{F} \to \mathbb{R}_{\geq 0}$ and $\rho \in \mathbb{R}_+$ at an alternative of the form

$$\mathscr{F}_{1,\rho} = \{f \in \mathscr{F} : \delta(f, f_o) \geq \rho\}.$$

The question arises how small the radius $\rho > 0$ can be chosen to obtain a good quality of testing, i.e., such that both of the above defined errors are small. Consequently, we measure the accuracy of a test Δ_n with its *maximal risk* defined by

$$\mathcal{R}\left(\Delta_n | \mathscr{F}_{1,\rho}\right) = \mathbb{P}^n_{f_o}\left(\Delta_n = 1\right) + \sup_{f \in \mathscr{F}_{1,\rho}} \mathbb{P}^n_f\left(\Delta_n = 0\right).$$

That is, we consider the sum of the type I error probability and the maximal type II probability over the alternative. Let $(\rho_n)_{n\in\mathbb{N}}$ be a sequence of values in $\mathbb{R}_+$ such that $\rho_n \to 0$ as $n \to \infty$. Then, a test statistic Δ_n is said to *attain the radius of testing* $(\rho_n)_{n\in\mathbb{N}}$ if for all $\alpha \in (0, 1)$ there exists a constant $A_\alpha \in \mathbb{R}_+$ such that for all $A \geq A_\alpha$ it holds

$$\mathcal{R}\left(\Delta_n | \mathscr{F}_{1,A\rho_n}\right) \leq \alpha.$$

Roughly speaking, this condition states that if we separate the null hypothesis and the alternative further than $A_\alpha \rho_n$, the maximal risk is smaller than α. There exists other notions of separation radii. We refer to Schluttenhofer (2020) for an overview.

Preferably, proposed tests perform optimally over a wide range of regularity classes simultaneously without prior knowledge of the underlying structure. Thus, in nonparametric hypothesis testing, adaptation is approached by multiple testing procedures. In this thesis we consider the Bonferroni method. Thus, we outline the main ideas of this method.

Let $\mathcal{K} \subseteq \mathbb{N}$ be a finite collection of dimension parameters. For a collection $(\Delta_k)_{k\in\mathcal{K}}$ of tests of level $\alpha/|\mathcal{K}| \in (0, 1)$, we consider the max-test

$$\Delta_{\mathcal{K},\alpha} := \mathbb{1}_{\{\zeta_{\mathcal{K},\alpha}>0\}} \quad \text{with} \quad \zeta_{\mathcal{K},\alpha} = \max_{k\in\mathcal{K}} \Delta_k,$$

i.e., the test rejects the null hypothesis as soon as one of the tests in the collection does. Under the null hypothesis, we bound the type I error probability of the max-test by the sum of the error probabilities of the individual tests

$$\mathbb{P}^n_{\varphi,f_o}\left(\Delta_{\mathcal{K},\alpha} = 1\right) = \mathbb{P}^n_{\varphi,f_o}\left(\zeta_{\mathcal{K},\alpha} > 0\right) \leq \sum_{k\in\mathcal{K}} \mathbb{P}^n_{\varphi,f_o}(\Delta_k = 1) \leq \sum_{k\in\mathcal{K}} \frac{\alpha}{|\mathcal{K}|} = \alpha.$$

Hence, $\Delta_{\mathcal{K},\alpha}$ is a level-α-test. Under the alternative the type II error probability is bounded by the minimum of the error probabilities of the individual tests, i.e.

$$\mathbb{P}^n_{\varphi,f}\left(\Delta_{\mathcal{K},\alpha} = 0\right) = \mathbb{P}^n_{\varphi,f}\left(\zeta_{\mathcal{K},\alpha} \leq 0\right) \leq \min_{k\in\mathcal{K}} \mathbb{P}^n_{\varphi,f}(\Delta_k = 0).$$

Therefore, $\Delta_{\mathcal{K},\alpha}$ has the maximal power achievable by a test in the collection. The bounds have opposing effects on the choice of the collection $\mathcal{K}$. On the one hand, it should be as small as possible to keep the type I error probability small. On the other hand, it must be large enough to approximate an optimal dimension parameter for a wide range of regularity assumptions that we want to adapt to.

Let us now introduce the specific statistical experiments considered in this thesis.

1.4 Convolution models

In this thesis, we consider two different convolution models. In Chapter 2 we consider the estimation of a transformation $\theta(f)$ of density f given by the additive convolution of marginal densities for direct observations. Consequently, the convolution is part of the parameter of interest. Conversely, in Chapters 3 and 4 we consider the case that observations are given from a transformation of the data $T(f)$ given by a multiplicative convolution with an error density. Let us introduce both models in this section.

1.4.1 Estimation of the p-fold additive convolution under direct observations

For $p \in \mathbb{N}$, we assume that $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(p)}$ are independent random variables in $(\mathbb{R}^d, \mathscr{B}^d, \lambda^d)$ admitting densities $f_1, \dots, f_p$, respectively. Let $\mathscr{F}$ denote the set of densities $(f_1, \dots, f_p)$ where each component defines a density on $(\mathbb{R}^d, \mathscr{B}^d, \lambda^d)$ additionally belonging to $\mathbb{L}^2_d$. Then, we consider the transformation $\theta\colon \mathscr{F} \to \mathbb{L}^2_d$, given for $f = (f_1, \dots, f_p) \in \mathscr{F}$ by the additive convolution

$$\theta(f) := f_1 * \dots * f_p.$$

We denote the parameter in the following by $g^{(p)} := \theta(f)$. Recall that the p-fold additive convolution of the component densities is well-defined and given for $\boldsymbol{x} \in \mathbb{R}^d$ by

$$f_1 * \dots * f_p(\boldsymbol{x}) = \int_{\mathbb{R}^d} \dots \int_{\mathbb{R}^d} f_1(\boldsymbol{x} - \boldsymbol{x}_2 - \dots - \boldsymbol{x}_p) f_2(\boldsymbol{x}_2) \dots f_p(\boldsymbol{x}_p) d\lambda^d(\boldsymbol{x}_2) \dots d\lambda^d(\boldsymbol{x}_p).$$

In Chapter 2 we consider the estimation of $g^{(p)}$ based on an independent set of i.i.d. samples $(\boldsymbol{X}_j^{(k)})_{j \in [\![n_k]\!]}$ for possibly different sample sizes $n_k \in \mathbb{N}$ and $k \in [\![p]\!]$. That is, we assume to have direct observations and the challenge lies in the parameter of interested given by the transformation θ. Recall that, applying recursively the additive convolution theorem given in Property 1.1.2, it follows that

$$\mathcal{F}[g^{(p)}] = \mathcal{F}[f_1] \cdots \mathcal{F}[f_p],$$

where $\mathcal{F}[g^{(p)}]$ denotes the Fourier transform of $g^{(p)}$ introduced in Section 1.1. In Chapter 2, we use this property to build two types of estimators, a kernel and a projection estimator, and analyze their behavior. Note that $g^{(p)}$ defines the density of the sum of components, i.e.,

$$\boldsymbol{X}^{(1)} + \dots + \boldsymbol{X}^{(p)} \sim g^{(p)}.$$

Consequently, $g^{(p)}$ can be interpreted as the density of aggregated data, which appears in many applications such as financial sciences. We give a more detailed motivation for this model in the literature review of Chapter 2.

1.4.2 Quadratic functional estimation and goodness-of-fit testing in the multiplicative measurement error model

For the multiplicative measurement error model, we consider the measure space $(\mathbb{R}_+, \mathscr{B}_+, \lambda_+)$. Let $\mathscr{F} \subseteq \mathbb{L}^2_+(\mathrm{x}^{2c-1}) \cap \mathbb{L}^1_+(\mathrm{x}^{c-1})$ and assume $X \sim f \in \mathscr{F}$. We assume to have access to observations Y admitting a density g given by a transformation $T(f)$ of f. More precisely, we consider the case that $T(f)$ is given by a the multiplicative convolution of f with some error density φ, i.e., for $y \in \mathbb{R}_+$

$$g(y) = T(f)(y) := (f \circledast \varphi)(y) = \int_{\mathbb{R}_+} f(x) \varphi(y/x) x^{-1} d\lambda_+(x). \tag{1.7}$$

Further, let U be a random variable independent of X admitting density φ. Then, we also describe the *multiplicative measurement error model* by

$$Y = X \cdot U.$$

Assuming that the error density φ is known, we have access to an i.i.d. sample $(Y_j)_{j\in[\![n]\!]}$ drawn from Y. As the name already suggests, this model is motivated by the fact that in practice observations might be contaminated by some measurement error. For a more detailed motivation, we refer to the literature review of Chapters 3 and 4.

Assuming throughout this work that the error density φ is known, we have access to an i.i.d. sample $(Y_j)_{j\in[\![n]\!]}$ drawn from Y. Applying the multiplicative convolution theorem given in Property 1.2.2, we obtain

$$\mathcal{M}_c[g] = \mathcal{M}_c[f] \cdot \mathcal{M}_c[\varphi].$$

This property plays a crucial role in the statistical analysis of Chapters 3 and 4. To specify the inference tasks of those chapters, we first define the following quadratic functional for functions $f \in \mathbb{L}^2_+(\mathrm{x}^{2c-1}) \cap \mathbb{L}^1_+(\mathrm{x}^{c-1})$ for fixed value $c \in \mathbb{R}$. Given an arbitrary measurable symmetric density function $\mathrm{w}\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ we define

$$q^2(f) := \|\mathcal{M}_c[f]\|^2_{\mathbb{L}^2(\mathrm{w}^2)} = \int_{\mathbb{R}} |\mathcal{M}_c[f](t)|^2\, \mathrm{w}^2(t) d\lambda(t), \tag{1.8}$$

where $\mathcal{M}_c[f]$ denotes the Mellin transform of f introduced in Section 1.4.2.

In Chapter 3, we consider the estimation of parameter $\theta(f) = q^2(f)$ given by the quadratic functional defined in Equation (1.8).

In Chapter 4, for a given density $f_o \in \mathscr{F}$ we consider the goodness-of-fit testing task introduced in Section 1.3.2. That is, we consider the null hypothesis

$$H_0\colon f = f_o.$$

The performance of a test is measured by how well it is able to distinguish between the null hypothesis in terms of the distance induced by the norm defined in Equation (1.8).

The following examples illustrate the range of quadratic functionals included in the definition of $q^2(f)$.

1.4.3 Examples of quadratic functionals

Depending on the explicit choice of the density function w and the value c, the general quadratic functional $q^2(f)$ defined in Equation (1.8) includes the following possible functionals of interest.

Example 1.4.1 (Quadratic functional of the density)**:**
If $\mathrm{w} = 1$, using the Parseval-type identity given in Property 1.2.1, we get that

$$q^2(f) = \|\mathcal{M}_c[f]\|^2_{\mathbb{L}^2} = \|f\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})} = \int_{\mathbb{R}_+} |f(x)|^2 x^{2c-1} d\lambda_+(x).$$

Consequently, we cover a quadratic functional evaluated at the function f itself. Note that, in the special case of $c = \frac{1}{2}$, it holds that $q^2(f) = \|f\|^2_{\mathbb{L}^2_+}$ and the quadratic functional is equal to the $\mathbb{L}^2_+$-norm of density f.

Example 1.4.2 (Quadratic functional of the survival function)**:**
If $\mathrm{w}^2(t) = \frac{1}{(c-1)^2+4\pi^2t^2}$ for all $t \in \mathbb{R}$, the parameter $q^2(f)$ equals the value of a quadratic functional evaluated at the survival function S_f of X. More precisely, for any density h, its survival function is given by

$$S_h : \mathbb{R}_+ \to \mathbb{R}_{\geq 0}, y \mapsto \int_y^\infty h(x) d\lambda_+(x).$$

Further, for any $c > 0$ it holds that $S_h \in \mathbb{L}^1_+(\mathrm{x}^{c-2})$ if and only if $h \in \mathbb{L}^1_+(\mathrm{x}^{c-1})$. For any $t \in \mathbb{R}$ it holds

$$\mathcal{M}_c[h](t) = (c - 1 + 2\pi i t)\mathcal{M}_{c-1}[S_h](t).$$

Thus, for $c > 0$ and $f \in \mathbb{L}^1_+(\mathrm{x}^{c-1})$ with $\mathcal{M}_c[f] \in \mathbb{L}^2(\mathrm{w}^2)$ it follows that

$$\begin{aligned} q^2(f) &= \int_{\mathbb{R}} \frac{|\mathcal{M}_c[f](t)|^2}{(c-1)^2 + 4\pi^2 t^2} d\lambda(t) = \int_{\mathbb{R}} \frac{|\mathcal{M}_{c-1}[S_f](t)|^2 |c - 1 + 2\pi i t|^2}{(c-1)^2 + 4\pi^2 t^2} d\lambda(t) \\ &= \|\mathcal{M}_{c-1}[S_f]\|^2_{\mathbb{L}^2} = \|S_f\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-3})}, \end{aligned}$$

using in the last step the Parseval-type identity given in Property 1.2.1. In the special case of $c = \frac{5}{2}$ the $\mathbb{L}^2_+$-norm $\|S_f\|^2_{\mathbb{L}^2_+}$ of the survival function S_f is considered. For more details and proofs for the mentioned properties of the survival functions see Brenner Miguel and Phandoidaen (2022).

Example 1.4.3 (Quadratic functional of derivatives of the density of interest)**:**
Let $f \in \mathcal{C}_0^\infty(\mathbb{R}_+)$, i.e., f is continuous with compact support in $\mathbb{R}_+$ such that for any $\beta \in \mathbb{N}$ the derivatives $D^\beta[f] := \frac{d^\beta}{dx^\beta} f$ exist. By Proposition 2.10, Brenner Miguel (2023) we have for $c \in \mathbb{R}$

$$\mathcal{M}_c[f](t) = (-1)^{-\beta} \frac{\Gamma(c + 2\pi i t)}{\Gamma(c + \beta + 2\pi i t)} \mathcal{M}_{c+\beta}[D^\beta[f]](t).$$

In addition, for $\mathrm{w}^2(t) = \prod_{j=1}^\beta ((c + \beta - j)^2 + 4\pi^2 t^2)$ for $\beta \in \mathbb{N}$ we have that

$$\begin{aligned} q^2(f) &= \int_{\mathbb{R}} |\mathcal{M}_c[f](t)|^2 \prod_{j=1}^{\beta} ((c + \beta - j)^2 + 4\pi^2 t^2) d\lambda(t) \\ &= \int_{\mathbb{R}} \left| \frac{\Gamma(c + 2\pi i t)}{\Gamma(c + \beta + 2\pi i t)} \mathcal{M}_{c+\beta}[D^\beta[f]](t) \right|^2 \prod_{j=1}^{\beta} ((c + \beta - j)^2 + 4\pi^2 t^2) d\lambda(t) \\ &= \|\mathcal{M}_{c+\beta}[D^\beta[f]]\|^2_{\mathbb{L}^2} = \|D^\beta[f]\|_{\mathbb{L}^2_+(\mathrm{x}^{2(c+\beta)-1})}, \end{aligned}$$

using that $\frac{\Gamma(c+2\pi i t)}{\Gamma(c+\beta+2\pi i t)} = \prod_{j=1}^\beta (c + \beta - j + 2\pi i t)$ and the Parseval-type identity given in Property 1.2.1. In the special case $c = \frac{1}{2} - \beta$ the $\mathbb{L}^2_+$-norm $\|D^\beta[f]\|^2_{\mathbb{L}^2_+}$ of the derivative $D^\beta[f]$ is considered.

This concludes this section. In the next chapter, we consider nonparametric estimation for the model introduced in Section 1.4.1.

CHAPTER 2

Estimation for the additive convolution of several multivariate densities

In this chapter, we consider the nonparametric estimation for the additive convolution of several multivariate densities.

More precisely, for $p \in \mathbb{N}$ we assume that $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(p)}$ are independent random vectors in $(\mathbb{R}^d, \mathscr{B}^d, \lambda^d)$. Further, for $j \in [\![p]\!]$ let component $\boldsymbol{X}^{(j)}$ admit a density $f_j \colon \mathbb{R}^d \to \mathbb{R}$ with respect to the Lebesgue measure λ^d belonging to $\mathbb{L}^2_d$. We consider the problem of estimating the density $g^{(p)}$ of the sum of components, that is,

$$\sum_{j \in [\![p]\!]} \boldsymbol{X}^{(j)} \sim g^{(p)}.$$

In this case, the density $g^{(p)}$ is equal to the the p-fold additive convolution of the densities f_j; in other words, we study the estimation of $g^{(p)} \colon \mathbb{R}^d \to \mathbb{R}$ defined for $\boldsymbol{x} \in \mathbb{R}^d$ by

$$\begin{aligned} g^{(p)}(\boldsymbol{x}) &:= f_1 * \cdots * f_p(\boldsymbol{x}) \\ &= \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} f_1(\boldsymbol{x} - \boldsymbol{x}_2 - \ldots - \boldsymbol{x}_p) f_2(\boldsymbol{x}_2) \ldots f_p(\boldsymbol{x}_p) d\lambda^d(\boldsymbol{x}_2) \ldots d\lambda^d(\boldsymbol{x}_p). \end{aligned} \tag{2.1}$$

That is, the challenge of this chapter is the estimation of a transformation $(f_1, \ldots, f_p) \mapsto f_1 * \cdots * f_p$ of the densities.

For the estimation, we assume to have access to $p \in \mathbb{N}$ many i.i.d. samples of independent vectors $\boldsymbol{X}^{(j)} \sim f_j$ with possibly different sample sizes $n_j \in \mathbb{N}$ for all $j \in [\![p]\!]$, i.e., we have access to observations

$$\boldsymbol{X}_k^{(j)} = (X_{k,1}^{(j)}, \ldots, X_{k,d}^{(j)}) \in \mathbb{R}^d, \quad k \in [\![n_j]\!], j \in [\![p]\!].$$

Throughout this chapter, we write shortly convolution referring to the additive convolution. The sum of independent random variables play a role in many fields. For example, in actuarial sciences the distribution of aggregate claims or ruin probabilities are of interest, see Panjer and Willmot (1992). For an extensive overview of further fields of application see Midhu et al. (2023).

Furthermore, in the estimation of compound Poisson processes, it is possible to relate its density to the convolution of the density of increments. In this context, estimators for convolutions of densities are then of interest; see Duval (2013). Comte et al. (2014) precisely plug in estimates of convolutions of any order of a density in order to build their specific Poisson-density estimator.

Related literature For certain families of distributions, the convolution of densities results in a change of parameters. For example, the sum of independent normally distributed random variables is again normally distributed. Other examples include the Gamma distribution (Hu et al. (2019)) and, in particular, exponentially distributed random variables, which are of interest in insurance risk theory (Willmot and Woo (2007)). Consequently, when estimating the density of the sum of independent variables merely leads to the estimation of adjusted parameters. In the case when the convolution of densities can not be derived analytically or the assumptions on the densities are not in a parametric form, a nonparametric estimator is useful.

For nonparametric estimation, one approach is to consider the special case of observing one real-valued random variable $X \sim f$ and estimating the p-fold convolution of f based on i.i.d. observations $(X_j)_{j \in [\![n]\!]}$ of X, $n \in \mathbb{N}$. This special case of the observation scheme discussed in this chapter has been considered in literature, however, for the estimation of a more general functional and nonparametric estimators of kernel type were proposed. More precisely, Frees (1994) considers the problem of estimating the density of functional a $q(X_1, \dots, X_m)$. The work was motivated by the measure of the degree of spatial randomness in a point pattern with $q(x_1, x_2) = |x_1 - x_2|$ but the results hold for a broader class of functions q, including the sum of observations. The proposed estimator of the density evaluated at a fixed point is based on a U-statistic and is proven to be point-wise $\sqrt{n}$-consistent. This estimator has also been extensively studied as random functions, see e.g. Ahmad and Mugdadi (2003) and Bandyopadhyay (2012). Giné and Mason (2007a,b) prove functional central limit theorems in various function spaces and provide a law of iterated logarithm for the density estimator. Ahmad and Fan (2001) deal with practical aspects of the choice of optimal bandwidths. Prakasa Rao (2004) derives moment inequalities for the deviation of the estimator from the true density. Further results on convergence in law for this estimator are given in a series of works by Schick and Wefelmeyer (2004b, 2008, 2009a,b, 2012).

As an alternative to Frees (1994), Saavedra and Cao (2000) consider for the special case of $q(X_1, X_2) = X_1 + aX_2$ a plug-in kernel estimator for its density, which can be written as the convolution of two densities; they obtain $\sqrt{n}$-consistency and an expression for mean squared and integrated mean squared error. Further, they extend their estimator to $\sum_{j \in [\![m]\!]} a_j X_j$ which may also cover the estimation of an m-fold convolution of a density. The estimator introduced in Saavedra and Cao (2000) has also been extensively studied, e.g., Schick and Wefelmeyer (2004a) prove asymptotic normality and efficiency. In Schick and Wefelmeyer (2007) it is shown that the estimators proposed by Saavedra and Cao (2000) and Frees (1994) are asymptotically equivalent, consequently, a series of results hold for both estimators.

Specifically for the m-fold convolution of a density, Chesneau et al. (2013) consider estimation relying on Fourier methods, they propose an adaptive estimator based on kernel methods and prove that bandwidth selection by Lepski method offers a general squared bias-variance compromise. Nickl (2007) obtains as a by-product of general results on nonparametric MLE estimators the $\sqrt{n}$-convergence in law of the convolution of two such estimators; see further Nickl (2009). Although previous work focused on a density on $\mathbb{R}$, these results can also be applied to the estimation of the convolution of two densities on $\mathbb{R}^d$.

An alternative to kernel estimation is to rely on projection methods. This was considered by Guo et al. (2018) for the m-fold convolution of a density; an estimate is defined as the m-fold convolution of a wavelet density estimator, and $\mathbb{L}^p$-consistency is shown.

We would like to mention that Midhu et al. (2023) consider the case of independent random variables possibly not identically distributed. They approximate the distribution function of the sum of two independent nonnegative random variables using quantile-based representation. In addition, in the case of dependent variables, there is no formula equivalent to Equation (2.1) for the density of the sum of variables, and very different tools are required; e.g. Cherubini et al. (2011) use the concept of Copula-based convolution.

Our focus lies on estimating the density of the sum of independent but not necessarily identically distributed random variables taking values in $\mathbb{R}^d$.

Contribution In this work, we study two approaches for the estimation of the density of the sum of p different d-dimensional independent random variables.

First, we show the link between the convolution of kernel estimators in the spirit of the plug-in approach of Saavedra and Cao (2000) and the Fourier formulation used in Chesneau et al. (2013). Then, we extend this estimator to the case of possibly different densities in d-dimensions. We show upper bounds of the point-wise and integrated quadratic risk and discuss the resulting convergence rates.

Second, we propose a general projection estimator and prove an upper bound on the integrated quadratic. Relying on the Hermite basis as described in Belomestny et al. (2019), we also build an adaptive estimator leading to a parsimonious development in the basis. This is the first time general results for projection estimators in terms of risk, rates and model selection are proved, as far as we know.

Outline of this chapter The chapter is structured as follows. In Section 2.1 we consider the kernel estimator, starting with the case of the convolution of two independent univariate variables in Section 2.1.1 and extending the estimator and results to the convolution of the densities of p independent multivariate variables in Section 2.1.2. In addition, we discuss the convergence rates in Sections 2.1.3 and 2.1.4 and propose an adaptive procedure in Section 2.1.5. Then, we turn to the projection estimator in Section 2.2. We first consider the case of two independent univariate variables in Section 2.2.1 and then introduce the Hermite basis in Section 2.2.2 to give the generalization in Section 2.2.3. In Section 2.2.4, we propose an adaptive procedure and show an upper bound for its integrated quadratic risk. Finally, Section 2.3 illustrates the results through a simulation study. Auxiliary results can be found in Section 2.4. Let us first recapitulate some notation used throughout this chapter.

Notation We use multi-index notation relying on bold letters for variables in $\mathbb{R}^d$ for $d \in \mathbb{N}$, that is, $\boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$. Operations on vectors $\boldsymbol{x} = (x_1, \dots, x_d), \boldsymbol{y} = (y_1, \dots, y_d) \in \mathbb{R}^d$ are to be interpreted component-wise. For example,

$$\boldsymbol{x}\boldsymbol{y} = (x_1 y_1, \dots, x_d y_d);$$

and

$$\boldsymbol{x} \le \boldsymbol{y} \Leftrightarrow (x_k \le y_k, \forall k \in [\![d]\!]).$$

Taking the sum over $\boldsymbol{j} \in [\![\boldsymbol{m}]\!]$ for $\boldsymbol{m} \in \mathbb{N}^d$ is to be understood as the sum over $j_k \in [\![m_k]\!]$ for all $k \in [\![d]\!]$. Further, for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ we denote by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{k \in [\![d]\!]} x_k y_k$$

the scalar product on $\mathbb{R}^d$. Let $h_1, h_2 \colon \mathbb{R}^d \to \mathbb{R}$ be two functions. We define the additive convolution $h_1 * h_2$, as soon as it is well-defined, e.g., for h_1 and h_2 integrable on $\mathbb{R}^d$, for $\boldsymbol{x} \in \mathbb{R}^d$ by

$$h_1 * h_2(\boldsymbol{x}) = \int_{\mathbb{R}^d} h_1(\boldsymbol{x} - \boldsymbol{y}) h_2(\boldsymbol{y}) d\lambda^d(\boldsymbol{y}).$$

The Fourier transform of an integrable function $h \colon \mathbb{R} \to \mathbb{R}$ is, for $\boldsymbol{t} \in \mathbb{R}^d$, given by

$$\mathcal{F}[h](\boldsymbol{t}) := \int_{\mathbb{R}^d} e^{i\langle \boldsymbol{t}, \boldsymbol{x} \rangle} h(\boldsymbol{x}) d\lambda^d(\boldsymbol{x}).$$

See Section 1.1 for more details and properties of the Fourier transform. For $p \in [1, \infty)$, the $\mathbb{L}^p_d$-norm of a function $h_1 \in \mathbb{L}^p_d$ is denoted by

$$\|h_1\|^2_{\mathbb{L}^p_d} := \int_{\mathbb{R}^d} |h_1(\boldsymbol{x})|^p d\lambda^d(\boldsymbol{x})$$

In addition, the essential supremum of a bounded function h is denoted by

$$\|h\|_{\mathbb{L}^\infty_d} := \sup_{\boldsymbol{x} \in \mathbb{R}^d} \{\lambda^d(\{\boldsymbol{y} \in \mathbb{R}^d : |h(\boldsymbol{y})| > \boldsymbol{x}\}) > 0\}.$$

We omit the index d in the case of $d = 1$, e.g., write $\mathbb{L}^p := \mathbb{L}^p_1$. Throughout this chapter we denote the expectation and probability measure with respect to any random variable X with $\mathbb{P}$ and $\mathbb{E}$, respectively, omitting any indices.

2.1 Kernel estimator

In this section, we propose a kernel estimator for the estimation of the convolution of two one-dimensional densities in a first step in Section 2.1.1, and several multivariate densities in a second step in Section 2.1.2. To extend the study to the case of the m-fold convolution, we will give another equivalent formula for the estimator and use Fourier strategies. We show upper bounds for its quadratic risk, consider convergence rates and propose a bandwidth estimator.

2.1.1 Convolution of the densities of two independent univariate variables

First, we consider the case of two independent one-dimensional variables to explain the estimation problem of interest and the proposed estimation method. In this case, we do not use the Fourier transform to derive an upper bound for the quadratic risks.

Throughout this section, let $X^{(1)}$ and $X^{(2)}$ be independent real-valued random variables admitting Lebesgue densities f_1 and f_2, respectively. Then, the sum $X^{(1)} + X^{(2)}$ also admits a Lebesgue density which we denote in this section by g. First note that $g\colon \mathbb{R} \to \mathbb{R}$ is equal to the (additive) convolution of the densities f_1 and f_2, that is, for $x \in \mathbb{R}$

$$g(x) = f_1 * f_2(x) := \int_{\mathbb{R}} f_1(y) f_2(x-y) d\lambda(y).$$

Given two sets of independent i.i.d. observations $(X_k^{(1)})_{k\in[\![n]\!]}$ of $X^{(1)}$ and $(X_k^{(2)})_{k\in[\![n]\!]}$ of $X^{(2)}$ for a common sample size $n \in \mathbb{N}$, the goal of this subsection is to propose a Kernel estimate of g. A natural proposal to estimate $g = f_1 * f_2$ is a plug-in estimator $\widehat{f}_{1,h} * \widehat{f}_{2,h}$, where $\widehat{f}_{1,h}$ and $\widehat{f}_{2,h}$ are kernel estimators for f_1 and f_2 given observations $(X_k^{(1)})_{k\in[\![n]\!]}$ and $(X_k^{(2)})_{k\in[\![n]\!]}$, respectively. More precisely, given an even kernel function K and a bandwidth $h \in \mathbb{R}_+$ we set for $j \in \{1, 2\}$

$$\widehat{f}_{j,h}(x) := \frac{1}{nh} \sum_{k\in[\![n]\!]} K\left(\frac{X_k^{(j)} - x}{h}\right), \quad x \in \mathbb{R}.$$

Definition 2.1.1:
A function $K\colon \mathbb{R} \to \mathbb{R}$ is called kernel if it satisfies $K \in \mathbb{L}^1$ with $\int_{\mathbb{R}} K(t) d\lambda(t) = 1$. If in addition it holds $K(-t) = K(t)$ for all $t \in \mathbb{R}$, K is said to be even.

Letting $L(x) := K * K(x)$ for $x \in \mathbb{R}$, the plug-in estimator $\widehat{f}_{1,h} * \widehat{f}_{2,h}$ can be rewritten for all $x \in \mathbb{R}$ as

$$\begin{aligned}
\widehat{f}_{1,h} * \widehat{f}_{2,h}(x) &= \frac{1}{n^2h^2} \sum_{k_1\in[\![n]\!]} \sum_{k_2\in[\![n]\!]} \int_{\mathbb{R}} K\left(\frac{X_{k_1}^{(1)} - (x-y)}{h}\right) K\left(\frac{X_{k_2}^{(2)} - y}{h}\right) d\lambda(y) \\
&= \frac{1}{n^2h} \sum_{k_1\in[\![n]\!]} \sum_{k_2\in[\![n]\!]} \int_{\mathbb{R}} K\left(\frac{X_{k_1}^{(1)} + X_{k_2}^{(2)} - x}{h} - u\right) K(u)\, d\lambda(u) \\
&= \frac{1}{n^2h} \sum_{k_1\in[\![n]\!]} \sum_{k_2\in[\![n]\!]} L\left(\frac{X_{k_1}^{(1)} + X_{k_2}^{(2)} - x}{h}\right). \qquad (2.2)
\end{aligned}$$

First, we show that L is again an even kernel.

Lemma 2.1.2:
The convolution of an even kernel K with itself $L := K * K$ is again an even kernel.

Proof of Lemma 2.1.2. We have for all $x \in \mathbb{R}$ that

$$\begin{aligned}
\int_{\mathbb{R}} |L(x)| d\lambda(x) &= \int_{\mathbb{R}} \left| \int_{\mathbb{R}} K(x-u) K(u) d\lambda(u) \right| d\lambda(x) \\
&\leq \int_{\mathbb{R}} |K(u)| \int_{\mathbb{R}} |K(x-u)| d\lambda(x) d\lambda(u) \\
&\leq \left| \int_{\mathbb{R}} |K(x)| d\lambda(x) \right|^2 < \infty.
\end{aligned}$$

Consequently, $L \in \mathbb{L}^1$. Moreover, it holds that

$$\begin{aligned}\int_{\mathbb{R}} L(x)d\lambda(x) &= \int_{\mathbb{R}}\int_{\mathbb{R}} K(x-u)K(u)d\lambda(u)d\lambda(x)\\ &= \int_{\mathbb{R}} K(u)\int_{\mathbb{R}} K(x-u)d\lambda(x)d\lambda(u)\\ &= \int_{\mathbb{R}} K(u)d\lambda(u)\int_{\mathbb{R}} K(y)d\lambda(y) = 1.\end{aligned}$$

This shows that L is indeed a kernel function. Further, it holds for $x \in \mathbb{R}$ that

$$\begin{aligned}L(x) &= \int_{\mathbb{R}} K(x-u)K(u)d\lambda(u) = \int_{\mathbb{R}} K(-x+u)K(u)d\lambda(u)\\ &= \int_{\mathbb{R}} K(-x-u)K(-u)d\lambda(u) = \int_{\mathbb{R}} K(-x-u)K(u)d\lambda(u) = L(-x),\end{aligned}$$

which shows that L is even and completes the proof. □

Consequently, the representation Equation (2.2) can be understood as a kernel estimator and we propose an estimator, where L is not necessarily a convolution of kernels but any arbitrary even kernel. More precisely, for an even kernel L, $h \in \mathbb{R}_+$ denote

$$L_h(x) := (1/h)L(x/h)$$

and define for $x \in \mathbb{R}$ the estimator

$$\widehat{g_h}(x) := \frac{1}{n^2}\sum_{k_1\in[\![n]\!]}\sum_{k_2\in[\![n]\!]} L_h(X_{k_1}^{(1)} + X_{k_2}^{(2)} - x), \tag{2.3}$$

This estimator can be seen as a kernel estimator from observations $(X_{k_1}^{(1)} + X_{k_2}^{(2)})_{k_1\in[\![n]\!],k_2\in[\![n]\!]}$ with kernel L, however, its summands are not independent. Let us remark some properties of this estimator. First, denoting $g_h := L_h * g$ it holds for any $x \in \mathbb{R}$ that

$$\begin{aligned}\mathbb{E}[\widehat{g_h}(x)] &= \mathbb{E}[L_h(X_{k_1}^{(1)} + X_{k_2}^{(2)} - x)]\\ &= \int_{\mathbb{R}} L_h(y-x)g(y)d\lambda(y) = g_h(x).\end{aligned} \tag{2.4}$$

Assuming $f_1, f_2 \in \mathbb{L}^2$, it follows that $g \in \mathbb{L}^2$. Moreover, if $L_h \in \mathbb{L}^2$, we also have $g_h \in \mathbb{L}^2$ and, in particular, $\widehat{g_h} \in \mathbb{L}^2$. Thus, the integrated quadratic risk of the estimator is well defined. Relying only on representation Equation (2.3), we prove the following result.

Proposition 2.1.3 (Bound on the integrated quadratic risk)**:**
If $f_1, f_2 \in \mathbb{L}^2$ and $L \in \mathbb{L}^2$ is an even kernel, then the estimator $\widehat{g_h}$ of g defined by (2.3) satisfies

$$\mathbb{E}\left[\|\widehat{g_h} - g\|_{\mathbb{L}^2}^2\right] \leq \|g_h - g\|_{\mathbb{L}^2}^2 + \frac{\|L\|_{\mathbb{L}^2}^2}{n^2 h} + \|L\|_{\mathbb{L}^1}^2 \frac{\|f_1\|_{\mathbb{L}^2}^2 + \|f_2\|_{\mathbb{L}^2}^2}{n},$$

where $g_h := L_h * g$.

Before giving a proof of Proposition 2.1.3, we shortly discuss the result and prove an auxiliary result central to its proof.

Remark 2.1.4:
Proposition 2.1.3 provides a squared-bias variance decomposition: the upper bound consists of the squared bias $\|g_h - g\|_{\mathbb{L}^2}^2$ and a variance term of order $1/(n^2 h) + 1/n$. The first part of the variance term results from the variance when summands are independent. The second term arises since the summands in the estimator are not independent, but does not depend on the bandwidth h and therefore has the role of a residual. If we had considered the n sample $(X_k^{(1)} + X_k^{(2)})_{k \in [\![n]\!]}$, we would have preserved the independence between summands and would have directly applied the standard kernel estimation result, see Comte (2017), chap.3; however, this way, we get a variance bound of order $1/(nh)$ instead of $1/(n^2 h) + 1/n$.

To bound the variance of the estimator Equation (2.3) for the convolution of two densities, we need to calculate the variance of a double sum. The next Lemma provides a decomposition of the variance, which will be also used for subsequent results.

We first give the Lemma and its proof before continuing with the proof of Proposition 2.1.3.

Lemma 2.1.5:
For $F \in \mathbb{L}^2$ we have that

$$\operatorname{var}\left(\frac{1}{n^2} \sum_{j \in [\![n]\!]} \sum_{k \in [\![n]\!]} F(X_j^{(1)} + X_k^{(2)})\right) = T_1 + T_2 + T_3,$$

with

$$\begin{aligned}
T_1 &:= \frac{1}{n} \frac{n-1}{n} \operatorname{cov}\left(F(X_1^{(1)} + X_1^{(2)}), F(X_1^{(1)} + X_2^{(2)})\right), \\
T_2 &:= \frac{1}{n} \frac{n-1}{n} \operatorname{cov}\left(F(X_1^{(1)} + X_1^{(2)}), F(X_2^{(1)} + X_1^{(2)})\right), \\
T_3 &:= \frac{1}{n^2} \operatorname{var}(F(X_1^{(1)} + X_1^{(2)})).
\end{aligned}$$

Proof of Lemma 2.1.5. Considering the variance, we split the sums to get

$$\begin{aligned}
&\operatorname{var}\left(\frac{1}{n^2} \sum_{j \in [\![n]\!]} \sum_{k \in [\![n]\!]} F(X_j^{(1)} + X_k^{(2)})\right) \\
&= \frac{1}{n^4} \sum_{j_1, j_2 \in [\![n]\!]} \sum_{k_1, k_2 \in [\![n]\!]} \operatorname{cov}\left(F(X_{j_1}^{(1)} + X_{k_1}^{(2)}), F(X_{j_2}^{(1)} + X_{k_2}^{(2)})\right).
\end{aligned}$$

Due to the independence assumption on the observations, we have that

$$\sum_{\substack{j_1, j_2 \in [\![n]\!] \\ j_1 \neq j_2}} \sum_{\substack{k_1, k_2 \in [\![n]\!] \\ k_1 \neq k_2}} \operatorname{cov}\left(F(X_{j_1}^{(1)} + X_{k_1}^{(2)}), F(X_{j_2}^{(1)} + X_{k_2}^{(2)})\right) = 0.$$

Splitting the sums appropriately we further get that

$$\begin{aligned}
&\operatorname{var}\Big(\frac{1}{n^2}\sum_{j\in[\![n]\!]}\sum_{k\in[\![n]\!]} F(X_j^{(1)}+X_k^{(2)})\Big)\\
&=\frac{1}{n^4}\Big(\sum_{j\in[\![n]\!]}\sum_{\substack{k_1,k_2\in[\![n]\!]\\ k_1\neq k_2}} \operatorname{cov}\Big(F(X_j^{(1)}+X_{k_1}^{(2)}),F(X_j^{(1)}+X_{k_2}^{(2)})\Big)\\
&\qquad+\sum_{k\in[\![n]\!]}\sum_{\substack{j_1,j_2\in[\![n]\!]\\ j_1\neq j_2}} \operatorname{cov}\Big(F(X_{j_1}^{(1)}+X_k^{(2)}),F(X_{j_2}^{(1)}+X_k^{(2)})\Big)\\
&\qquad+\sum_{j\in[\![n]\!]}\sum_{k\in[\![n]\!]} \operatorname{var}(F(X_j^{(1)}+X_k^{(2)}))\Big).
\end{aligned}$$

Further, due to the i.i.d. assumption on the observations,for each line summands are equal and the result follows

$$\operatorname{var}\Big(\frac{1}{n^2}\sum_{j\in[\![n]\!]}\sum_{k\in[\![n]\!]} F(X_j^{(1)}+X_k^{(2)})\Big)=T_1+T_2+T_3.$$

This concludes the proof. □

We return to the proof of Proposition 2.1.3 applying the last Lemma with an appropriate choice of F.

Proof of Proposition 2.1.3. Since for any $x\in\mathbb{R}$ it holds that $\mathbb{E}[\widehat{g}_h(x)]=g_h(x)$, see Equation (2.4), with the standard decomposition it holds that

$$\mathbb{E}\left[\|\widehat{g}_h-g\|_{\mathbb{L}^2}^2\right]=\|g_h-g\|_{\mathbb{L}^2}^2+\mathbb{E}\left[\|\widehat{g}_h-g_h\|_{\mathbb{L}^2}^2\right].$$

It remains to upper bound the second term. Applying Fubini's theorem, we have that

$$\begin{aligned}
\mathbb{E}\left[\|\widehat{g}_h-g_h\|_{\mathbb{L}^2}^2\right]&=\mathbb{E}\left[\int_{\mathbb{R}}\left|\frac{1}{n^2}\sum_{k_1\in[\![n]\!]}\sum_{k_2\in[\![n]\!]} L_h\left(X_{k_1}^{(1)}+X_{k_2}^{(2)}-x\right)-L_h*g(x)\right|^2 d\lambda(x)\right]\\
&=\int_{\mathbb{R}}\operatorname{var}\left(\frac{1}{n^2}\sum_{k_1\in[\![n]\!]}\sum_{k_2\in[\![n]\!]} L_h\left(X_{k_1}^{(1)}+X_{k_2}^{(2)}-x\right)\right)d\lambda(x). \qquad (2.5)
\end{aligned}$$

Choosing $F(y)=L_h(y-x)$ we use Lemma 2.1.5, to decompose the variance and, therefore, the integral in three terms. First, note that $L\in\mathbb{L}^2$ by assumption and, thus, $F\in\mathbb{L}^2$ for any $x\in\mathbb{R}$, as well. We consider first T_3. For $x\in\mathbb{R}$ we have that

$$T_3=\frac{1}{n^2}\operatorname{var}(L_h(X_1^{(1)}+X_1^{(2)}-x)).$$

The corresponding term of the integral Equation (2.5) can be upper bounded by

$$\begin{aligned}
\int_{\mathbb{R}}\frac{1}{n^2}\operatorname{var}(L_h(X_1^{(1)}+X_1^{(2)}-x))d\lambda(x)&\leq\frac{1}{n^2h^2}\int_{\mathbb{R}}\int_{\mathbb{R}} g(y)\left|L\left(\frac{y-x}{h}\right)\right|^2 d\lambda(y)d\lambda(x)\\
&=\frac{1}{n^2h}\int_{\mathbb{R}} g(y)\lambda(y)\int_{\mathbb{R}}|L(x)|^2d\lambda(x)=\frac{\|L\|_{\mathbb{L}^2}^2}{n^2h}. \qquad (2.6)
\end{aligned}$$

Considering T_1, we first see that for $x \in \mathbb{R}$

$$\begin{aligned}
\frac{n^2}{n-1} T_1 &= \operatorname{cov}\left(L_h(X_1^{(1)} + X_1^{(2)} - x), L_h(X_1^{(1)} + X_2^{(2)} - x)\right) \\
&= \mathbb{E}\left[L_h(X_1^{(1)} + X_1^{(2)} - x) L_h(X_1^{(1)} + X_2^{(2)} - x)\right] - \mathbb{E}\left[L_h(X_1^{(1)} + X_1^{(2)} - x)\right]^2 \\
&\leq \mathbb{E}\left[L_h(X_1^{(1)} + X_1^{(2)} - x) L_h(X_1^{(1)} + X_2^{(2)} - x)\right].
\end{aligned}$$

Recall that $X_1^{(1)}$ and $X_1^{(2)}, X_2^{(2)}$ are independent. Further, $X_1^{(2)}$ and $X_2^{(2)}$ are independent and identically distributed. It follows that

$$\begin{aligned}
&\mathbb{E}\left[L_h(X_1^{(1)} + X_1^{(2)} - x) L_h(X_1^{(1)} + X_2^{(2)} - x)\right] \\
&= \int_{\mathbb{R}^d} \mathbb{E}\left[L_h(u + X_1^{(2)} - x) L_h(u + X_2^{(2)} - x)\right] f_1(u) d\lambda^d(u) \\
&= \int_{\mathbb{R}^d} \mathbb{E}\left[L_h(u + X_1^{(2)} - x)\right]^2 f_1(u) d\lambda^d(u) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} L_h(u + v - x) f_2(v) d\lambda(v)\right)^2 f_1(u) d\lambda(u)
\end{aligned}$$

It follows that the term corresponding to T_1 of the integral in Equation (2.5) is upper bounded by

$$\begin{aligned}
&\frac{n-1}{n^2} \int_{\mathbb{R}} \operatorname{cov}\left(L_h(X_1^{(1)} + X_1^{(2)} - x), L_h(X_1^{(1)} + X_2^{(2)} - x)\right) d\lambda(x) \\
&\leq \frac{1}{n} \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} L_h(u + v - x) f_2(v) d\lambda(v)\right)^2 f_1(u) d\lambda(u) d\lambda(x) \qquad (2.7)
\end{aligned}$$

Young's inequality, see Property 1.1.3 (ii), for L_h and f_2 yields

$$\|L_h * f_2\|_{\mathbb{L}^2}^2 \leq \|L_h\|_{\mathbb{L}^1}^2 \|f_2\|_{\mathbb{L}^2}^2.$$

Using additionally the definition of the convolution and that f_1 is a density, it follows that

$$\begin{aligned}
&\frac{1}{n} \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} L_h(u + v - x) f_2(v) d\lambda(v)\right)^2 f_1(u) d\lambda(u) d\lambda(x) \\
&= \frac{1}{n} \int_{\mathbb{R}} f_1(u) \int_{\mathbb{R}} \left(L_h * f_2(x - u)\right)^2 d\lambda(x) d\lambda(u) \\
&= \frac{1}{n} \int_{\mathbb{R}} f_1(u) \int_{\mathbb{R}} \left(L_h * f_2(x)\right)^2 d\lambda(x) d\lambda(u) \\
&\leq \frac{1}{n} \|L_h\|_{\mathbb{L}^1}^2 \|f_2\|_{\mathbb{L}^2}^2 = \frac{1}{n} \|L\|_{\mathbb{L}^1}^2 \|f_2\|_{\mathbb{L}^2}^2. \qquad (2.8)
\end{aligned}$$

Due to the symmetry of T_1 and T_2, we get for the integral term of Equation (2.5) corresponding to T_2 the upper bound

$$\frac{1}{n} \|L\|_{\mathbb{L}^1}^2 \|f_1\|_{\mathbb{L}^2}^2. \qquad (2.9)$$

Finally, combining Equations (2.6), (2.8) and (2.9) to upper bound Equation (2.5) yields

$$\mathbb{E}\left[\|\widehat{g}_h - g_h\|_{\mathbb{L}^2}^2\right] \leq \frac{\|L\|_{\mathbb{L}^2}^2}{n^2 h} + \|L\|_{\mathbb{L}^1}^2 \frac{\|f_1\|_{\mathbb{L}^2}^2 + \|f_2\|_{\mathbb{L}^2}^2}{n},$$

which concludes the proof. □

We note that a bounded density f also belongs to $\mathbb{L}^2_d$ since

$$\int_{\mathbb{R}} |f(t)|^2 d\lambda(t) \le \|f\|_{\mathbb{L}^\infty} \int_{\mathbb{R}} f(t) d\lambda(t) = \|f\|_{\mathbb{L}^\infty} < \infty.$$

Consequently, if $f_1, f_2 \in \mathbb{L}^\infty$ it follows $f_1, f_2 \in \mathbb{L}^2$ and, thus, $g = f_1 * f_2 \in \mathbb{L}^2$. Additionally, it follows that the Fourier transform $\mathcal{F}[g]$ of g is in $\mathbb{L}^1$. More precisely, with Cauchy–Schwarz inequality it follows

$$\begin{aligned}\int_{\mathbb{R}} |\mathcal{F}[g](x)| d\lambda(x) &= \int_{\mathbb{R}} |\mathcal{F}[f_1](x)\mathcal{F}[f_2](x)| d\lambda(x) \\ &\le \left(\int_{\mathbb{R}} |\mathcal{F}[f_1](x)|^2 d\lambda(x)\right)^{1/2} \left(\int_{\mathbb{R}} |\mathcal{F}[f_1](x)|^2 d\lambda(x)\right)^{1/2} \\ &= \|f_1\|_{\mathbb{L}^2} \|f_2\|_{\mathbb{L}^2} < \infty.\end{aligned}$$

Finally, it also follows that $g \in \mathbb{L}^\infty$ since for any $x \in \mathbb{R}$

$$\begin{aligned}|g(x)| = |\mathcal{F}^\dagger[\mathcal{F}[g]](x)| &= \left| \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ixt} \mathcal{F}[g](t) d\lambda(t) \right| \\ &\le \frac{1}{2\pi} \int_{\mathbb{R}} |\mathcal{F}[g]t| d\lambda(t) = \frac{1}{2\pi} \|\mathcal{F}[g]\|_{\mathbb{L}^1} < \infty.\end{aligned}$$

Consequently, in this case, point evaluation of g is well-defined and we also provide a point-wise risk bound.

Proposition 2.1.6 (Bound on the point-wise quadratic risk)**:**
If $f_1, f_2 \in \mathbb{L}^\infty$ and $L \in \mathbb{L}^2$ is an even kernel, then the estimator $\widehat{g}_h$ of g defined by (2.3) satisfies for $x \in \mathbb{R}$ that

$$\mathbb{E}\left[(\widehat{g}_h(x) - g(x))^2\right] \le (g_h(x) - g(x))^2 + \|L\|^2_{\mathbb{L}^1} \frac{\|g\|_{\mathbb{L}^\infty}}{n^2 h} + \frac{C_f \|L\|^2_{\mathbb{L}^1}}{n},$$

where $C_f := \|f_1\|_{\mathbb{L}^\infty} \|f_2\|^2_{\mathbb{L}^2} + \|f_2\|_{\mathbb{L}^\infty} \|f_1\|^2_{\mathbb{L}^2}$.

Remark 2.1.7:
The upper bound on the point-wise quadratic risk consists of the squared bias $(g_h(x) - g(x))^2$ and and a variance term of order $1/(n^2 h) + 1/n$. The first part of the variance term results from the variance when summands are independent. The second term arises since the summands in the estimator are not independent, but does not depend on the bandwidth h and therefore has the role of a residual. Proposition 2.1.6 improves the bound of Proposition 1 in Chesneau et al. (2013) for the case $m = 2$, who obtain a variance term of order $1/(n^2 h^2)$ instead of $1/(n^2 h)$ here. To get two independent samples, as are considered in their work, given an i.i.d. $(X_k)_{k \in [\![n]\!]}$ sample one could split the sample in two and consequently get two independent samples of size of $n/2$ each. The point of Remark 2.1.4 also holds for the point-wise risk.

Proof of Proposition 2.1.6. Since for any $x \in \mathbb{R}$ it holds $\mathbb{E}[\widehat{g}_h(x)] = g_h(x)$, see Equation (2.4), we have the standard decomposition

$$\mathbb{E}\left[(\widehat{g}_h(x) - g(x))^2\right] = \mathbb{E}\left[(\widehat{g}_h(x) - g_h(x))^2\right] + (g_h(x) - g(x))^2.$$

Analogously to Equation (2.5) in the proof of Proposition 2.1.3 it holds that

$$\mathbb{E}\left[(\widehat{g}_h(x) - g_h(x))^2\right] = \operatorname{var}\left(\frac{1}{n^2}\sum_{k_1\in[\![n]\!]}\sum_{k_2\in[\![n]\!]} L_h\left(X_{k_1}^{(1)} + X_{k_2}^{(2)} - x\right)\right).$$

Therefore, we use again Lemma 2.1.5 with $F(y) = L_h(y - x)$ and split the variance into three terms. By assumption $L \in \mathbb{L}^2$ and, thus, $F \in \mathbb{L}^2$ for any $x \in \mathbb{R}$. We get for T_3 that

$$T_3 = \frac{1}{n^2}\operatorname{var}(L_h(X_1^{(1)} + X_1^{(2)} - x)) \le \frac{1}{n^2h^2}\int_{\mathbb{R}} g(y)\left|L\left(\frac{y-x}{h}\right)\right|^2 d\lambda(y) \le \frac{1}{n^2h}\|g\|_{\mathbb{L}^\infty}\|L\|_{\mathbb{L}^2}^2.$$

For T_1 with the calculations in Equation (2.7) and using that L_h is an even kernel, we get

$$\begin{aligned}
T_1 =& \frac{n-1}{n^2}\operatorname{cov}\left(L_h(X_1^{(1)} + X_1^{(2)} - x), L_h(X_1^{(1)} + X_2^{(2)} - x)\right)\\
\le& \frac{1}{n}\int_{\mathbb{R}}\left(\int_{\mathbb{R}} L_h(u+v-x) f_2(v) d\lambda(v)\right)^2 f_1(u) d\lambda(u)\\
\le& \frac{1}{n}\|f_1\|_{\mathbb{L}^\infty}\int_{\mathbb{R}}\left(\int_{\mathbb{R}} L_h(u+v-x) f_2(v) d\lambda(v)\right)^2 d\lambda(u)\\
=& \frac{1}{n}\|f_1\|_{\mathbb{L}^\infty}\int_{\mathbb{R}}\left(\int_{\mathbb{R}} L_h(u-v) f_2(v) d\lambda(v)\right)^2 d\lambda(u)
\end{aligned}$$

Applying again Young's inequality, see Property 1.1.3 (ii), for L_h and f_2 yields

$$T_1 \le \frac{1}{n}\|f_1\|_{\mathbb{L}^\infty}\|L_h * f_2\|_{\mathbb{L}^2}^2 \le \frac{1}{n}\|f_1\|_{\mathbb{L}^\infty}\|L\|_{\mathbb{L}^1}^2\|f_2\|_{\mathbb{L}^2}^2$$

Analogously, T_2 can be upper bounded by $\frac{1}{n}\|f_2\|_{\mathbb{L}^\infty}\|L\|_{\mathbb{L}^1}^2\|f_1\|_{\mathbb{L}^2}^2$. Consequently, combing all three bounds, it holds that

$$\mathbb{E}\left[(\widehat{g}_h(x) - g_h(x))^2\right] \le \frac{\|g\|_{\mathbb{L}^\infty}\|L\|_{\mathbb{L}^2}^2}{n^2h} + \|L\|_{\mathbb{L}^1}^2\frac{\|f_1\|_{\mathbb{L}^\infty}\|f_2\|_{\mathbb{L}^2}^2 + \|f_2\|_{\mathbb{L}^\infty}\|f_1\|_{\mathbb{L}^2}^2}{n}$$

which concludes the proof. □

2.1.2 Convolution of the densities of p-many independent multivariate variables

In this section we generalize the results from Section 2.1.1 to the convolution of multiple multivariate densities. Now, the goal is to estimate $g^{(p)}\colon \mathbb{R}^d \to \mathbb{R}$ defined by Equation (2.1) from independent vector observations which are all d-dimensional but with possibly different sample sizes, i.e., for $j \in [\![p]\!]$, we observe $\boldsymbol{X}_k^{(j)} = (X_{k,1}^{(j)}, \dots, X_{k,d}^{(j)}) \in \mathbb{R}^d$, with $\boldsymbol{X}_k^{(j)} \sim f_j$ i.i.d., for $k \in [\![n_j]\!]$ and independent for $j \in [\![p]\!]$.

To propose a generalized kernel estimator, let us extend Definition 2.1.1 of an even function.

Definition 2.1.8:
A function $L\colon \mathbb{R}^d \to \mathbb{R}$ is called even if $L(-\boldsymbol{t}) = L(\boldsymbol{t})$ holds for all $\boldsymbol{t} \in \mathbb{R}^d$.

Since in this section we use different analytical tools based on the Fourier transform, we no further need to assume $L \in \mathbb{L}_d^1$. That is, L is not assumed to be a kernel function in the sense of Definition 2.1.1. However, the general form of the estimator is the same and we still refer to the method as a kernel estimator. Further, to apply properties of the Fourier transform additional assumptions are need, which we specify below in Assumption 2.1.10.

For $L\colon \mathbb{R}^d \to \mathbb{R}$ and $\boldsymbol{h} = (h_1, \dots, h_d) \in \mathbb{R}_+^d$ we denote

$$L_{\boldsymbol{h}}(\boldsymbol{x}) := \frac{1}{h_1 \dots h_d} L\left(\frac{x_1}{h_1}, \dots, \frac{x_d}{h_d}\right). \tag{2.10}$$

Then, for an even function $L\colon \mathbb{R}^d \to \mathbb{R}$, bandwidth $\boldsymbol{h} = (h_1, \dots, h_d) \in \mathbb{R}_+^d$ and sample size $\boldsymbol{n} = (n_1, \dots, n_p) \in \mathbb{N}^p$ the estimator $\widehat{g_{\boldsymbol{h}}^{(p)}}$ of $g^{(p)}$ at $\boldsymbol{x} \in \mathbb{R}^d$ is defined as

$$\widehat{g_{\boldsymbol{h}}^{(p)}}(\boldsymbol{x}) := \frac{1}{n_1 \cdots n_p} \sum_{\boldsymbol{k} \in [\![\boldsymbol{n}]\!]} L_{\boldsymbol{h}}\left(\boldsymbol{X}_{k_1}^{(1)} + \dots + \boldsymbol{X}_{k_p}^{(p)} - \boldsymbol{x}\right). \tag{2.11}$$

This extends the estimator from Equation (2.3) on the one hand to the convolution of p different densities and on the other hand to the d-dimensional case using independent samples of different sizes. Nevertheless, our direct computations to prove Proposition 2.1.3 would be tedious to extend. More precisely, dealing with the decomposition of the variance is complicated for products of more than two elements. Thus, we use Fourier methods and, to that aim, we rewrite the estimator in the Fourier domain. Recall that for $j \in [\![p]\!]$ we denote by $\mathcal{F}[f_j]$ the Fourier transform of density f_j as introduced in Section 1.1, and define its empirical counterpart by

$$\widehat{\mathcal{F}[f_j]}(\boldsymbol{t}) := \frac{1}{n_j} \sum_{k \in [\![n_j]\!]} e^{i\langle \boldsymbol{t}, \boldsymbol{X}_k^{(j)} \rangle}, \quad \boldsymbol{t} \in \mathbb{R}^d. \tag{2.12}$$

Then, $\widehat{g_{\boldsymbol{h}}^{(p)}}$ defined by Equation (2.11) can also be written in terms of the Fourier transform.

Lemma 2.1.9:
If $L \in \mathbb{L}_d^2$ is an even function and with integrable Fourier transform $\mathcal{F}[L] \in \mathbb{L}_d^1$, it holds for the estimator defined in Equation (2.11) that for $\boldsymbol{x} \in \mathbb{R}^d$

$$\widehat{g_{\boldsymbol{h}}^{(p)}}(\boldsymbol{x}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle \boldsymbol{t}, \boldsymbol{x} \rangle} \mathcal{F}[L](\boldsymbol{h}\boldsymbol{t}) \prod_{j \in [\![p]\!]} \widehat{\mathcal{F}[f_j]}(\boldsymbol{t}) d\lambda^d(\boldsymbol{t}). \tag{2.13}$$

Proof of Lemma 2.1.9. Since $L \in \mathbb{L}_d^2$, we rewrite Equation (2.11) taking the Fourier inverse of its Fourier transform, i.e.,

$$\begin{aligned}
&\widehat{g_{\boldsymbol{h}}^{(p)}}(\boldsymbol{x}) \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle \boldsymbol{u}, \boldsymbol{x} \rangle} \mathcal{F}[\widehat{g_{\boldsymbol{h}}^{(p)}}](\boldsymbol{u}) d\lambda^d(\boldsymbol{u}) \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle \boldsymbol{u}, \boldsymbol{x} \rangle} \frac{1}{n_1 \cdots n_p} \sum_{\boldsymbol{j} \in [\![\boldsymbol{n}]\!]} \int_{\mathbb{R}^d} e^{i\langle \boldsymbol{u}, \boldsymbol{y} \rangle} L_{\boldsymbol{h}}\left(\boldsymbol{X}_{j_1}^{(1)} + \dots + \boldsymbol{X}_{j_p}^{(p)} - \boldsymbol{y}\right) d\lambda^d(\boldsymbol{y})\, d\lambda^d(\boldsymbol{u}).
\end{aligned}$$

By change of variables and the definition of the empirical Fourier transform of each density $\widehat{\mathcal{F}[f_j]}$ defined in Equation (2.12) it further follows that

$$\begin{aligned}
&\widehat{g_{\boldsymbol{h}}^{(p)}}(\boldsymbol{x}) \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}} e^{-i\langle \boldsymbol{u},\boldsymbol{x}\rangle} \frac{1}{n_1 \cdots n_p} \sum_{\boldsymbol{j}\in[\![\boldsymbol{n}]\!]} \int_{\mathbb{R}^d} L(\boldsymbol{w}) e^{i\langle \boldsymbol{u}, \boldsymbol{X}_{j_1}^{(1)} + \ldots + \boldsymbol{X}_{j_p}^{(p)} - \boldsymbol{h}\boldsymbol{w}\rangle} d\lambda^d(\boldsymbol{w})\, d\lambda^d(\boldsymbol{u}) \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle \boldsymbol{u},\boldsymbol{x}\rangle} \mathcal{F}[L](\boldsymbol{h}\boldsymbol{u}) \prod_{j\in[\![p]\!]} \widehat{\mathcal{F}[f_j]}(\boldsymbol{u})\, d\lambda^d(\boldsymbol{u}).
\end{aligned}$$

This completes the proof. □

Let us relate the estimator $\widehat{g_{\boldsymbol{h}}^{(p)}}$ to the estimator proposed in Chesneau et al. (2013) for the case of dimension $d = 1$. If we start from the formula given in Equation (2.13) and take L as the sinus cardinal kernel, i.e., for $x \in \mathbb{R}$,

$$L(x) := \frac{\sin(\pi x)}{\pi x} \tag{2.14}$$

we have that L admits a Fourier transform given by

$$\mathcal{F}[L](hu) = \mathbb{1}_{[-\pi,\pi]}(hu) = \mathbb{1}_{[-\pi/h,\pi/h]}(u)$$

and the estimator writes

$$\widehat{g_h^{(p)}}(x) = \frac{1}{2\pi} \int_{[-\frac{\pi}{h},\frac{\pi}{h}]} e^{-iux} \prod_{j\in[\![p]\!]} \widehat{\mathcal{F}[f_j]}(u)\, d\lambda(u).$$

This shows that Equation (2.13) extends the estimator (7) of Chesneau et al. (2013) in three directions: first, to the case of differing densities; second, for a general function L; third, to multivariate densities. Note that we use both kernel methods to analyze the estimator using Equation (2.11) and Fourier analysis using Equation (2.13). The assumptions required on L may differ, but with no main drawback, as it is user-chosen.

To proof upper bounds for the integrated and point-wise quadratic risk of estimator $\widehat{g_{\boldsymbol{h}}^{(p)}}$, we consider the following assumptions.

Assumption 2.1.10: (i) L is an even function which belongs to $\mathbb{L}^2_d$.

(ii) For the Fourier transform of L it holds $\mathcal{F}[L] \in \mathbb{L}^1_d \cap \mathbb{L}^\infty_d$.

Note that from $L \in \mathbb{L}^2_d$ it follows $\mathcal{F}[L] \in \mathbb{L}^2_d$. Let us remark that for L and $\boldsymbol{h} \in \mathbb{R}^d_+$ the Fourier transform of $L_{\boldsymbol{h}}$, defined in Equation (2.10), satisfies

$$\mathcal{F}[L_{\boldsymbol{h}}](\boldsymbol{x}) = \int_{\mathbb{R}^d} e^{i\langle \boldsymbol{x},\boldsymbol{t}\rangle} L_{\boldsymbol{h}}(\boldsymbol{t}) d\lambda^d(\boldsymbol{t}) = \int_{\mathbb{R}^d} e^{i\langle \boldsymbol{x},\boldsymbol{h}\boldsymbol{t}\rangle} L(\boldsymbol{t}) d\lambda^d(\boldsymbol{t}) = \mathcal{F}[L](\boldsymbol{h}\boldsymbol{x}).$$

Under Assumption 2.1.10, we define

$$g_{\boldsymbol{h}}^{(p)}(\boldsymbol{x}) := L_{\boldsymbol{h}} * g^{(p)}(\boldsymbol{x}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle \boldsymbol{t},\boldsymbol{x}\rangle} \mathcal{F}[L](\boldsymbol{h}\boldsymbol{t}) \prod_{j\in[\![p]\!]} \mathcal{F}[f_j](\boldsymbol{t})\, d\lambda^d(\boldsymbol{t}). \tag{2.15}$$

For the estimator defined in Equation (2.11) it holds that for all $\boldsymbol{x} \in \mathbb{R}^d$

$$\mathbb{E}\left[\widehat{g_{\boldsymbol{h}}^{(p)}}(\boldsymbol{x})\right] = \mathbb{E}\left[L_{\boldsymbol{h}}\left(\boldsymbol{X}_{k_1}^{(1)} + \ldots + \boldsymbol{X}_{k_p}^{(p)} - \boldsymbol{x}\right)\right]$$
$$= \int_{\mathbb{R}^d} L_{\boldsymbol{h}}(\boldsymbol{y} - \boldsymbol{x}) g^{(p)}(\boldsymbol{x}) d\lambda^d(\boldsymbol{y}) = L_{\boldsymbol{h}} * g^{(p)}(\boldsymbol{x}) = g_{\boldsymbol{h}}^{(p)}(\boldsymbol{x}). \tag{2.16}$$

In the following sections, we consider first the integrated quadratic risk for the proposed estimator and derive convergence rates in Section 2.1.3. Afterwards, we consider its point-wise quadratic risk and also derive convergence rates in Section 2.1.4.

2.1.3 Integrated quadratic risk

For the integrated quadratic risk, we show the following upper bound.

Proposition 2.1.11 (Bound on the integrated quadratic risk):
Let L satisfy Assumption 2.1.10 and $f_j \in \mathbb{L}_d^2$ for all $j \in [\![p]\!]$. Then, the estimator $\widehat{g_{\boldsymbol{h}}^{(p)}}$ of $g^{(p)}$ defined in Equation (2.13) satisfies

$$\mathbb{E}\left[\|\widehat{g_{\boldsymbol{h}}^{(p)}} - g^{(p)}\|_{\mathbb{L}_d^2}^2\right] \leq \|g_{\boldsymbol{h}}^{(p)} - g^{(p)}\|_{\mathbb{L}_d^2}^2 + \frac{2\|L\|_{\mathbb{L}_d^2}^2}{h_1 \cdots h_d n_1 \cdots n_p} + C \sum_{j \in [\![p]\!]} \frac{1}{n_j},$$

with $C := 2p3^{2(p-2)} \|\mathcal{F}[L]\|_{\mathbb{L}_d^\infty}^2 \max_{j \in [\![p]\!]} \|f_j\|_{\mathbb{L}_d^2}^2$.

Remark 2.1.12:
The upper bound on the integrated quadratic risk consists of the squared bias $\|g_{\boldsymbol{h}}^{(p)} - g^{(p)}\|_{\mathbb{L}_d^2}^2$, an extended variance of order $h_1 \cdots h_d n_1 \cdots n_p$ and a residual term of order $\max_{j \in [\![p]\!]} 1/n_j$ which does not depend on the bandwidth.

Note that if for some $j \in [\![p]\!]$ the sample size n_j is small the residual term causes the order residual term behaves correspondingly. If all sample sizes are of the same order $n_j = \alpha_j n$ for some $n \in \mathbb{N}$ and constants α_j, $j \in [\![p]\!]$, then, the variance term is simplified to order $1/(h_1 \cdots h_d n^p)$ and the residual term of order $1/n$. In addition, we see that this result is indeed a generalization of Proposition 2.1.3 in the sense that got setting $p = 2$ and $d = 1$ we obtain the same result.

Furthermore, if $L(\boldsymbol{x}) = \prod_{k \in [\![d]\!]} L_k(x_k)$, then $\mathcal{F}[L](\boldsymbol{t}) = \prod_{k \in [\![d]\!]} \mathcal{F}[L_k](t_k)$, and all L_k can be taken as the sinus cardinal kernel defined in Equation (2.14). In this case, writing for the Cartesian product of intervals $\mathcal{H} := \times_{k \in [\![d]\!]} [-\pi/h_k, \pi h_k]$, it holds

$$\|g_{\boldsymbol{h}}^{(p)} - g^{(p)}\|_{\mathbb{L}_d^2}^2 = \int_{\mathbb{R}^d \setminus \mathcal{H}} |\mathcal{F}[g^{(p)}](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}).$$

Thus, the bias term converges to zero if $\max_{k \in [\![d]\!]} h_k \to 0$ by dominated convergence.

Proof of Proposition 2.1.11. First, due to Equation (2.16), it holds that $\mathbb{E}\left[\widehat{g_{\boldsymbol{h}}^{(p)}}(\boldsymbol{x})\right] = g_{\boldsymbol{h}}^{(p)}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^d$ and we decompose the risk as usual

$$\mathbb{E}\left[\|\widehat{g_{\boldsymbol{h}}^{(p)}} - g^{(p)}\|_{\mathbb{L}_d^2}^2\right] = \|g_{\boldsymbol{h}}^{(p)} - g^{(p)}\|_{\mathbb{L}_d^2}^2 + \mathbb{E}\left[\|\widehat{g_{\boldsymbol{h}}^{(p)}} - g_{\boldsymbol{h}}^{(p)}\|_{\mathbb{L}_d^2}^2\right].$$

Having representation Equation (2.13) in mind, Parseval's identity, see Property 1.1.1, gives for the variance term

$$\mathbb{E}\left[\|\widehat{g_{\boldsymbol{h}}^{(p)}} - g_{\boldsymbol{h}}^{(p)}\|_{\mathbb{L}_d^2}^2\right] = \frac{1}{(2\pi)^d}\mathbb{E}\left[\int_{\mathbb{R}^d}\left|\mathcal{F}[L](\boldsymbol{ht})\prod_{j\in[\![p]\!]}\widehat{\mathcal{F}[f_j]}(\boldsymbol{t}) - \prod_{j\in[\![p]\!]}\mathcal{F}[f_j](\boldsymbol{t})\right|^2 d\lambda^d(\boldsymbol{t})\right]$$
$$= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}|\mathcal{F}[L](\boldsymbol{ht})|^2\mathbb{E}\left[\left|\prod_{j\in[\![p]\!]}\widehat{\mathcal{F}[f_j]}(\boldsymbol{t}) - \prod_{j\in[\![p]\!]}\mathcal{F}[f_j](\boldsymbol{t})\right|^2\right] d\lambda^d(\boldsymbol{t}). \tag{2.17}$$

For fixed $\boldsymbol{t} \in \mathbb{R}^d$ we apply Lemma 2.4.1, given below in Section 2.4, for the term inside the expectation. First, note that for all $j \in [\![p]\!]$ it holds $\widehat{\mathcal{F}[f_j]}(\boldsymbol{t}), \mathcal{F}[f_j](\boldsymbol{t}) \in \mathbb{C}$ with $|\widehat{\mathcal{F}[f_j]}(\boldsymbol{t})|, |\mathcal{F}[f_j](\boldsymbol{t})| \leq 1$. Consequently, we apply Lemma 2.4.1 with $u_j = \widehat{\mathcal{F}[f_j]}(\boldsymbol{t})$ and $v_j = \mathcal{F}[f_j](\boldsymbol{t})$ for $j \in [\![p]\!]$. We obtain

$$\left|\prod_{j\in[\![p]\!]}\widehat{\mathcal{F}[f_j]}(\boldsymbol{t}) - \prod_{j\in[\![p]\!]}\mathcal{F}[f_j](\boldsymbol{t})\right| \leq \prod_{j\in[\![p]\!]}|\widehat{\mathcal{F}[f_j]}(\boldsymbol{t}) - \mathcal{F}[f_j](\boldsymbol{t})|$$
$$+ \sum_{j\in[\![p]\!]} 3^{p-2}|\widehat{\mathcal{F}[f_j]}(\boldsymbol{t}) - \mathcal{F}[f_j](\boldsymbol{t})||\mathcal{F}[f_{j+1}](\boldsymbol{t})|$$

with the notation that $\mathcal{F}[f_{p+1}] = \mathcal{F}[f_1]$. Next, we take the expectation of the square of each side. Using the definition of the empirical Fourier transform, see Equation (2.12), and the independence between the random variables $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(p)}$ we get for $\boldsymbol{t} \in \mathbb{R}^d$ that

$$\mathbb{E}\left[\left|\prod_{j\in[\![p]\!]}\widehat{\mathcal{F}[f_j]}(\boldsymbol{t}) - \prod_{j\in[\![p]\!]}\mathcal{F}[f_j](\boldsymbol{t})\right|^2\right] \leq 2\prod_{j\in[\![p]\!]}\mathbb{E}\left[\left|\frac{1}{n_j}\sum_{k\in[\![n_j]\!]}e^{i\langle \boldsymbol{t}, \boldsymbol{X}_k^{(j)}\rangle} - \mathcal{F}[f_j](\boldsymbol{t})\right|^2\right]$$
$$+ 2p\cdot 3^{2(p-2)}\sum_{j\in[\![p]\!]}|\mathcal{F}[f_{j+1}](\boldsymbol{t})|^2\mathbb{E}\left[\left|\frac{1}{n_j}\sum_{k\in[\![n_j]\!]}e^{i\langle \boldsymbol{t}, \boldsymbol{X}_k^{(j)}\rangle} - \mathcal{F}[f_j](\boldsymbol{t})\right|^2\right].$$

For expectations, we see for $\boldsymbol{t} \in \mathbb{R}^d$ and $j \in [\![p]\!]$ that

$$\mathbb{E}\left[\left|\frac{1}{n_j}\sum_{k\in[\![n_j]\!]}e^{i\langle \boldsymbol{t}, \boldsymbol{X}_k^{(j)}\rangle} - \mathcal{F}[f_j](\boldsymbol{t})\right|^2\right] = \frac{1}{n_j}\operatorname{var}\left(e^{i\langle \boldsymbol{t}, \boldsymbol{X}_k^{(j)}\rangle}\right)$$
$$\leq \frac{1}{n_j}\mathbb{E}\left[|e^{i\langle \boldsymbol{t}, \boldsymbol{X}_k^{(j)}\rangle}|\right] \leq \frac{1}{n_j}. \tag{2.18}$$

With this follows

$$\mathbb{E}\left[\left|\prod_{j\in[\![p]\!]}\left(\frac{1}{n_j}\sum_{k\in[\![n_j]\!]}e^{i\langle \boldsymbol{t}, \boldsymbol{X}_k^{(j)}\rangle}\right) - \prod_{j\in[\![p]\!]}\mathcal{F}[f_j](\boldsymbol{t})\right|^2\right] \leq \frac{2}{n_1\cdots n_p} + 2p\cdot 3^{2(p-2)}\sum_{j\in[\![p]\!]}\frac{|\mathcal{F}[f_{j+1}](\boldsymbol{t})|^2}{n_j}. \tag{2.19}$$

Plugging the last inequality into Equation (2.17), we get

$$\mathbb{E}\left[\|\widehat{g_{\boldsymbol{h}}^{(p)}} - g_{\boldsymbol{h}}^{(p)}\|_{\mathbb{L}_d^2}^2\right] \leq \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} |\mathcal{F}[L](\boldsymbol{ht})|^2 \Big(\frac{2}{n_1\cdots n_p} + 2p\cdot 3^{2(p-2)} \sum_{j\in[\![p]\!]} \frac{|\mathcal{F}[f_{j+1}](\boldsymbol{t})|^2}{n_j}\Big) d\lambda^d(\boldsymbol{t})$$
$$= \frac{1}{(2\pi)^d}\frac{2}{n_1\cdots n_p}\int_{\mathbb{R}^d} |\mathcal{F}[L](\boldsymbol{ht})|^2 d\lambda^d(\boldsymbol{t})$$
$$+ \frac{2p\cdot 3^{2(p-2)}}{(2\pi)^d}\sum_{j\in[\![p]\!]}\frac{1}{n_j}\int_{\mathbb{R}^d} |\mathcal{F}[L](\boldsymbol{ht})|^2 |\mathcal{F}[f_{j+1}](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}). \tag{2.20}$$

For the integral of the first summand we use Parseval's identity, see Property 1.1.1, and see that

$$\int_{\mathbb{R}^d} |\mathcal{F}[L](\boldsymbol{ht})|^2 d\lambda^d(\boldsymbol{t}) = \frac{1}{h_1\cdots h_d}\|\mathcal{F}[L]\|_{\mathbb{L}_d^2} = \frac{(2\pi)^d}{h_1\cdots h_d}\|L\|_{\mathbb{L}_d^2}.$$

For the integral of the second summand we use again Parseval's identity and get for any $j \in [\![p]\!]$ that

$$\int_{\mathbb{R}^d} |\mathcal{F}[L](\boldsymbol{ht})\mathcal{F}[f_j](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}) \leq \|\mathcal{F}[L]\|_{\mathbb{L}_d^\infty}^2 \|\mathcal{F}[f_j]\|_{\mathbb{L}_d^2}^2 \leq (2\pi)^d \|L\|_{\mathbb{L}_d^1}^2 \|f_j\|_{\mathbb{L}_d^2}^2$$
$$\leq (2\pi)^d \|\mathcal{F}[L]\|_{\mathbb{L}_d^\infty}^2 \max_{j\in[\![p]\!]} \|f_j\|_{\mathbb{L}_d^2}^2.$$

Consequently, inserting the last two bounds into Equation (2.20) we obtain the result

$$\mathbb{E}\left[\|\widehat{g_{\boldsymbol{h}}^{(p)}} - g_{\boldsymbol{h}}^{(p)}\|_{\mathbb{L}_d^2}^2\right] \leq \frac{2\|L\|_{\mathbb{L}_d^2}}{h_1\cdots h_d n_1\cdots n_p} + C\sum_{j\in[\![p]\!]}\frac{1}{n_j},$$

with $C := 2p3^{2(p-2)}\|\mathcal{F}[L]\|_{\mathbb{L}_d^\infty}^2 \max_{j\in[\![p]\!]} \|f_j\|_{\mathbb{L}_d^2}^2$. This shows the result and concludes the proof. □

Convergence rates

Let us consider convergence rates for the kernel estimator defined in Equation (2.13) based on its integrated quadratic risk. More precisely, we discuss the order of the upper bound given in Proposition 2.1.11. For this, we consider the case of anisotropic Sobolev spaces with an additional boundedness assumption generalizing the one proposed in Chesneau et al. (2013) and the case of Nikolski classes. We state the results by imposing the regularity on $g^{(p)}$. First, we give a definition of both classes for $d \in \mathbb{N}$ dimensions. Subsequently, we give a link between the regularity of $g^{(p)}$ and its component densities f_j, $j \in [\![p]\!]$.

Definition 2.1.13 (Anisotropic Sobolev space)**:**
A function $h\colon \mathbb{R}^d \to \mathbb{R}_+$ belongs to Sobolev space $\mathcal{S}(\boldsymbol{\alpha}, R)$ for $\alpha \in \mathbb{R}_+^d$ and $R \in \mathbb{R}_+$ if $h \in \mathbb{L}_d^2$ and for its Fourier transform $\mathcal{F}[h]$ it holds that

$$\sum_{k\in[\![d]\!]}\int_{\mathbb{R}^d}(1+t_k^2)^{\alpha_k}|\mathcal{F}[h](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}) \leq R^2 \quad \text{and} \quad \max_{k\in[\![d]\!]}\sup_{\boldsymbol{t}\in\mathbb{R}^d}(1+t_k^2)^{\alpha_k}|\mathcal{F}[h](\boldsymbol{t})|^2 \leq R^2.$$

To be able to control the bias for anisotropic Sobolev classes, we need to ensure that the Fourier transform of L is a real-valued function. Further, we consider the following assumptions extending the assumptions introduced by Taupin (2001) for dimension $d = 1$.

Assumption 2.1.14:
The Fourier transform of L satisfies $\mathcal{F}[L](\boldsymbol{t}) = 1$ for any $\boldsymbol{t} \in [-1,1]^d$.

For dimension $d = 1$, the sinus cardinal kernel defined in Equation (2.14) satisfies Assumption 2.1.10 and Assumption 2.1.14, see Taupin (2001) Remark 2.2. Taupin (2001) points out that the de La Vallée-Poussin kernel V for $x \in \mathbb{R}$ defined by

$$V(x) = \frac{\cos(x) - \cos(2x)}{\pi x^2}$$

also satisfies Assumption 2.1.10 and Assumption 2.1.14. Next, we give a definition of Nikolski classes for functions on $\mathbb{R}^d$ as in Comte and Lacour (2013). For $a \in \mathbb{R}_{\geq 0}$ we denote by $\lfloor a \rfloor$ its integer part. For a function $h\colon \mathbb{R}^d \to \mathbb{R}$ that is differentiable of order $l \in \mathbb{N}$ in the k-th argument x_k, we write $\frac{\partial^l h}{(\partial x_k)^l}$ for the l-th derivative of h in argument x_k.

Definition 2.1.15 (Nikolski class):
We say that $h\colon \mathbb{R}^d \to \mathbb{R}_+$ belongs to Nikolski space $\mathcal{N}(\boldsymbol{\alpha}, R)$ for $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}_+^d$ and $R \in \mathbb{R}_+$, if f is up to order $\lfloor \alpha_k \rfloor$ times differentiable in x_k for all $k \in [\![d]\!]$ and

(i) $\| \frac{\partial^{\lfloor \alpha_k \rfloor} h}{(\partial x_k)^{\lfloor \alpha_k \rfloor}} \|_{\mathbb{L}_d^2} \leq R$ for all $k \in [\![d]\!]$;

(ii) For all $k \in [\![d]\!]$ for all $y \in \mathbb{R}$ it holds that

$$\left(\int_{\mathbb{R}^d} \left| \frac{\partial^{\lfloor \alpha_k \rfloor} h}{(\partial x_k)^{\lfloor \alpha_k \rfloor}} (x_1, \ldots, x_{k-1}, x_k + y, x_{k+1}, \ldots, x_d) - \frac{\partial^{\lfloor \alpha_k \rfloor} h}{(\partial x_k)^{\lfloor \alpha_k \rfloor}} (\boldsymbol{x}) \right|^2 d\lambda^d(\boldsymbol{x}) \right)^{1/2} \leq R |y|^{\alpha_k - \lfloor \alpha_k \rfloor}.$$

As is standard in density estimation, see Comte (2017), for $f \in \mathcal{N}(\boldsymbol{\alpha}, R)$ we need the additional assumption of the kernel L to be of a given order to handle the bias.

Definition 2.1.16:
An even kernel function $L\colon \mathbb{R}^d \to \mathbb{R}$ is of order $\boldsymbol{\alpha} \in \mathbb{N}^d$ if

$$\forall k \in [\![d]\!] : \forall 1 \leq l \leq \alpha_k : \int_{\mathbb{R}^d} x_k^l L(\boldsymbol{x}) d\lambda^d(\boldsymbol{x}) = 0,$$
$$\forall k \in [\![d]\!] : \int_{\mathbb{R}^d} (1 + |x_k|)^{\alpha_k} |L(\boldsymbol{x})| d\lambda^d(\boldsymbol{x}) < \infty.$$

Moreover, to control the bias for Nikolski classes, we need the additional assumption of L to be a kernel function, analogously to Definition 2.1.1.

Assumption 2.1.17:
Let $L \in \mathbb{L}_d^1$ with $\int_{\mathbb{R}^d} L(\boldsymbol{t}) d\lambda^d(\boldsymbol{t}) = 1$.

First, we determine the bias for both function classes.

Lemma 2.1.18:
Let Assumption 2.1.10 be satisfied, $\boldsymbol{\alpha} \in \mathbb{R}_+^d$ and $R \in \mathbb{R}_+$. If in addition one of the following conditions is satisfied:

(i) Assumption 2.1.14 holds and $g^{(p)} \in \mathcal{S}(\boldsymbol{\alpha}, R)$,

(ii) Assumption 2.1.17 holds and L is of order $(\lfloor\alpha_1\rfloor, \ldots, \lfloor\alpha_d\rfloor)$ and $g^{(p)} \in \mathcal{N}(\boldsymbol{\alpha}, R)$,

then, there exists a constant $C \in \mathbb{R}_+$ such that for $\boldsymbol{h} = (h_1, \ldots, h_d) \in \mathbb{R}_+^d$ and $g_{\boldsymbol{h}}^{(p)}$ defined in Equation (2.15) it holds that

$$\|g_{\boldsymbol{h}}^{(p)} - g^{(p)}\|_{\mathbb{L}_d^2}^2 \leq C(h_1^{2\alpha_1} + \ldots + h_d^{2\alpha_d}).$$

Proof of Lemma 2.1.18. The following computations are standard, see for example Tsybakov (2008). We give the proofs for the sake of completeness. First, consider (i). Define the set

$$\mathcal{H} := [-\frac{1}{h_1}, \frac{1}{h_1}] \times \cdots \times [-\frac{1}{h_d}, \frac{1}{h_d}].$$

On $\mathcal{H}$ we have by Assumption 2.1.14 that $1 - \mathcal{F}[L](\boldsymbol{h}\boldsymbol{t}) = 0$. Using additionally Parseval's identity, see Property 1.1.1, and the definition of the Sobolev space we get

$$\begin{aligned}
\|g_{\boldsymbol{h}}^{(p)} - g^{(p)}\|_{\mathbb{L}_d^2}^2 &= \int_{\mathbb{R}^d} |1 - \mathcal{F}[L](\boldsymbol{h}\boldsymbol{t})|^2 |\mathcal{F}[g^{(p)}](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}) \\
&= \int_{\mathbb{R}^d \setminus \mathcal{H}} |1 - \mathcal{F}[L](\boldsymbol{h}\boldsymbol{t})|^2 |\mathcal{F}[g^{(p)}](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}) \\
&\leq \frac{2 + 2\|\mathcal{F}[L]\|_{\mathbb{L}_d^\infty}^2}{d} \sum_{k \in [\![d]\!]} \int_{\mathbb{R}^d \setminus \mathcal{H}} (1 + t_k^2)^{\alpha_k} (1 + t_k^2)^{-\alpha_k} |\mathcal{F}[g^{(p)}](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}).
\end{aligned}$$

Note that on $\mathbb{R}^d \setminus \mathcal{H}$ it holds that $(1 + t_k^2)^{-\alpha_k} \leq h_k^{2\alpha_k}$. It follows

$$\begin{aligned}
\|g_{\boldsymbol{h}}^{(p)} - g^{(p)}\|_{\mathbb{L}_d^2}^2 &\leq \frac{2 + 2\|\mathcal{F}[L]\|_{\mathbb{L}_d^\infty}^2}{d} (h_1^{2\alpha_1} + \ldots + h_d^{2\alpha_d}) \sum_{k \in [\![d]\!]} \int_{\mathbb{R}^d} (1 + t_k^2)^{-\alpha_k} |\mathcal{F}[g^{(p)}](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}) \\
&\leq R^2 \frac{2 + 2\|\mathcal{F}[L]\|_{\mathbb{L}_d^\infty}^2}{d} (h_1^{2\alpha_1} + \ldots + h_d^{2\alpha_d}).
\end{aligned}$$

Consider (ii). Using that L integrates to one and is even, we get for $\boldsymbol{x} \in \mathbb{R}^d$ that

$$\begin{aligned}
g_{\boldsymbol{h}}^{(p)}(\boldsymbol{x}) - g^{(p)}(\boldsymbol{x}) = L_{\boldsymbol{h}} * g^{(p)}(\boldsymbol{x}) - g^{(p)}(\boldsymbol{x}) &= \int_{\mathbb{R}^d} L_{\boldsymbol{h}}(\boldsymbol{t} - \boldsymbol{x}) g^{(p)}(\boldsymbol{t}) d\lambda^d(\boldsymbol{t}) - g^{(p)}(\boldsymbol{x}) \\
&= \int_{\mathbb{R}^d} L(\boldsymbol{u}) g^{(p)}(\boldsymbol{h}\boldsymbol{u} + \boldsymbol{x}) d\lambda^d(\boldsymbol{u}) - \int_{\mathbb{R}^d} L(\boldsymbol{u}) g^{(p)}(\boldsymbol{x}) d\lambda^d(\boldsymbol{u}) \\
&= \int_{\mathbb{R}^d} L(\boldsymbol{u})(g^{(p)}(\boldsymbol{h}\boldsymbol{u} + \boldsymbol{x}) - g^{(p)}(\boldsymbol{x})) d\lambda^d(\boldsymbol{u}). \qquad (2.21)
\end{aligned}$$

Since $g^{(p)} \in \mathcal{N}(\boldsymbol{\alpha}, R)$ we apply Taylor-expansion component-wise and write for $\boldsymbol{u} \in \mathbb{R}^d$

$$
\begin{aligned}
&g^{(p)}(\boldsymbol{hu}+\boldsymbol{x}) - g^{(p)}(\boldsymbol{x}) \\
&= \sum_{j \in [\![d]\!]} \Bigg(\sum_{k=1}^{\lfloor \alpha_j \rfloor -1} \frac{(u_j h_j)^k}{k!} \frac{\partial^k (g^{(p)})}{(\partial x_j)^k}(x_1, \ldots, x_j, u_{j+1}h_{j+1}+x_{j+1}, \ldots, u_d h_d + x_d) \\
&\qquad + \frac{(u_j h_j)^{\lfloor \alpha_j \rfloor}}{(\lfloor \alpha_j \rfloor - 1)!} \int_0^1 (1-\tau)^{\lfloor \alpha_j \rfloor -1} \\
&\qquad\quad \cdot \frac{\partial^{\lfloor \alpha_j \rfloor}(g^{(p)})}{(\partial x_j)^{\lfloor \alpha_j \rfloor}}(x_1, \ldots, x_{j-1}, \tau u_j h_j + x_j, u_{j+1}h_{j+1}+x_{j+1}, \ldots, u_d h_d + x_d) d\lambda(\tau) \Bigg).
\end{aligned}
$$

Using that L is a kernel of oder $(\lfloor \alpha_1 \rfloor, \ldots \lfloor \alpha_d \rfloor)$ we get for $j \in [\![d]\!]$ and $k \le \lfloor \alpha_j \rfloor$ that

$$
0 = \int_{\mathbb{R}^d} L(\boldsymbol{u}) \frac{(u_j h_j)^k}{k!} \frac{\partial^k g^{(p)}}{(\partial x_j)^k}(x_1, \ldots, x_j, u_{j+1}h_{j+1}+x_{j+1}, \ldots, u_d h_d + x_d) d\lambda^d(\boldsymbol{u})
$$

With the last two calculations, it follows that

$$
\begin{aligned}
&\int_{\mathbb{R}^d} L(\boldsymbol{u})(g^{(p)}(\boldsymbol{hu}+\boldsymbol{x}) - g^{(p)}(\boldsymbol{x})) d\lambda^d(\boldsymbol{u}) \\
&= \int_{\mathbb{R}^d} \sum_{j \in [\![d]\!]} \frac{h_j^{\lfloor \alpha_j \rfloor}}{(\lfloor \alpha_j \rfloor - 1)!} L(\boldsymbol{u}) u_j^{\lfloor \alpha_j \rfloor} \int_0^1 (1-\tau)^{\lfloor \alpha_j \rfloor -1} \\
&\quad \cdot \Bigg(\frac{\partial^{\lfloor \alpha_j \rfloor} g^{(p)}}{(\partial x_j)^{\lfloor \alpha_j \rfloor}}(x_1, \ldots, x_{j-1}, \tau u_j h_j + x_j, u_{j+1}h_{j+1}+x_{j+1}, \ldots, u_d h_d + x_d) \\
&\qquad - \frac{\partial^{\lfloor \alpha_j \rfloor} g^{(p)}}{(\partial x_j)^{\lfloor \alpha_j \rfloor}}(x_1, \ldots, x_j, u_{j+1}h_{j+1}+x_{j+1}, \ldots, u_d h_d + x_d) \Bigg) \, d\lambda(\tau) \, d\lambda^d(\boldsymbol{u}).
\end{aligned}
$$

With this we get for the integrated quadratic risk that

$$
\begin{aligned}
&\|g_{\boldsymbol{h}}^{(p)} - g^{(p)}\|_{\mathbb{L}_d^2}^2 \\
&= \int_{\mathbb{R}^d} \Bigg[\int_{\mathbb{R}^d} \sum_{j \in [\![d]\!]} \frac{h_j^{\lfloor \alpha_j \rfloor}}{(\lfloor \alpha_j \rfloor - 1)!} L(\boldsymbol{u}) u_j^{\lfloor \alpha_j \rfloor} \int_0^1 (1-\tau)^{\lfloor \alpha_j \rfloor -1} \\
&\quad \cdot \Bigg(\frac{\partial^{\lfloor \alpha_j \rfloor} g^{(p)}}{(\partial x_j)^{\lfloor \alpha_j \rfloor}}(x_1, \ldots, x_{j-1}, \tau u_j h_j + x_j, u_{j+1}h_{j+1}+x_{j+1}, \ldots, u_d h_d + x_d) \\
&\qquad - \frac{\partial^{\lfloor \alpha_j \rfloor} g^{(p)}}{(\partial x_j)^{\lfloor \alpha_j \rfloor}}(x_1, \ldots, x_j, u_{j+1}h_{j+1}+x_{j+1}, \ldots, u_d h_d + x_d) \Bigg) \, d\lambda(\tau) \, d\lambda^d(\boldsymbol{u}) \Bigg]^2 \, d\lambda^d(\boldsymbol{x}).
\end{aligned}
$$

Further, it follows

$$
\begin{aligned}
&\|g_{\boldsymbol{h}}^{(p)} - g^{(p)}\|_{\mathbb{L}_d^2}^2 \\
&\leq d \sum_{j \in [\![d]\!]} \frac{h_j^{2\lfloor \alpha_j \rfloor}}{((\lfloor \alpha_j \rfloor - 1)!)^2} \int_{\mathbb{R}^d} \Bigg[\int_{\mathbb{R}^d} L(\boldsymbol{u}) u_j^{\lfloor \alpha_j \rfloor} \int_0^1 (1-\tau)^{\lfloor \alpha_j \rfloor - 1} \\
&\qquad \cdot \Bigg(\frac{\partial^{\lfloor \alpha_j \rfloor} g^{(p)}}{(\partial x_j)^{\lfloor \alpha_j \rfloor}} (x_1, \ldots, x_{j-1}, \tau u_j h_j + x_j, u_{j+1} h_{j+1} + x_{j+1}, \ldots, u_d h_d + x_d) \\
&\qquad\qquad - \frac{\partial^{\lfloor \alpha_j \rfloor} g^{(p)}}{(\partial x_j)^{\lfloor \alpha_j \rfloor}} (x_1, \ldots, x_j, u_{j+1} h_{j+1} + x_{j+1}, \ldots, u_d h_d + x_d) \Bigg) \, d\lambda(\tau) \, d\lambda^d(\boldsymbol{u}) \Bigg]^2 d\lambda^d(\boldsymbol{x}).
\end{aligned}
$$

Consequently, applying the Generalized Minkowski inequality, see Lemma 2.4.6, two times and using that $g^{(p)} \in \mathcal{N}(\boldsymbol{\alpha}, R)$ we get that

$$
\begin{aligned}
&\|g_{\boldsymbol{h}}^{(p)} - g^{(p)}\|_{\mathbb{L}_d^2}^2 \\
&\leq d \sum_{j \in [\![d]\!]} \frac{h_j^{2\lfloor \alpha_j \rfloor}}{((\lfloor \alpha_j \rfloor - 1)!)^2} \left[R'' \int_{\mathbb{R}^d} |L(\boldsymbol{u}) u_j^{\lfloor \alpha_j \rfloor}| \int_0^1 (1-\tau)^{\lfloor \alpha_j \rfloor - 1} |\tau u_j h_j|^{\alpha_j - \lfloor \alpha_j \rfloor} d\lambda(\tau) \, d\lambda^d(\boldsymbol{u}) \right]^2 \\
&\leq R^2 d \sum_{j \in [\![d]\!]} \frac{h_j^{2\alpha_j}}{((\lfloor \alpha_j \rfloor - 1)!)^2} \left(\int_{\mathbb{R}^d} |L(\boldsymbol{u}) u_j^{\alpha_j}| d\lambda^d(\boldsymbol{u}) \right)^2 \\
&\leq R^2 \max_{j \in [\![d]\!]} \left(\frac{1}{((\lfloor \alpha_j \rfloor - 1)!)^2} \int_{\mathbb{R}^d} |L(\boldsymbol{u}) u_j^{\alpha_j}| d\lambda^d(\boldsymbol{u}) \right)^2 d \sum_{j \in [\![d]\!]} h_j^{2\alpha_j}.
\end{aligned}
$$

This concludes the proof. □

Next, we discuss the convergence rates for the quadratic integrated risk. For this, we denote by $\overline{\alpha}$ the harmonic mean of the regularity components α_j, i.e.

$$
\frac{1}{\overline{\alpha}} = \frac{1}{d} \left(\frac{1}{\alpha_1} + \ldots + \frac{1}{\alpha_d} \right). \tag{2.22}
$$

Combining the last result for the bias, Lemma 2.1.18, with the upper bound of Proposition 2.1.11, we obtain the following rates of convergence for the estimator $\widehat{g_{\boldsymbol{h}}^{(p)}}$.

Proposition 2.1.19:
Let Assumption 2.1.10 be satisfied. Assume that all n_j are of the same order $n \in \mathbb{N}$, that is, $n_j = c_j n$ for c_j constants independent of n, for all $j \in [\![p]\!]$. In addition, assume that one the following conditions is satisfied for $\boldsymbol{\alpha} \in \mathbb{R}_+^d$ and $R \in \mathbb{R}_+$:

(i) Assumption 2.1.14 holds and $g^{(p)} \in \mathcal{S}(\boldsymbol{\alpha}, R)$,

(ii) Assumption 2.1.17 holds and L is of order $(\lfloor \alpha_1 \rfloor, \dots, \lfloor \alpha_d \rfloor)$ and $g^{(p)} \in \mathcal{N}(\boldsymbol{\alpha}, R)$.

Then, with $\boldsymbol{h}_{\text{opt}} \in \mathbb{R}_+^d$ defined via $h_{k,\text{opt}} := n^{-\frac{\overline{\alpha} p}{\alpha_k(2\overline{\alpha}+d)}}$ for $k \in [\![d]\!]$, there exists a constant $C \in \mathbb{R}_+$ such that

$$\mathbb{E}\left[\|\widehat{g_{\boldsymbol{h}_{\text{opt}}}^{(p)}} - g^{(p)}\|_{\mathbb{L}_d^2}^2\right] \leq C \begin{cases} n^{-1}, & \text{if } \overline{\alpha} \geq \frac{d}{2(p-1)}, \\ n^{-\frac{2\overline{\alpha}p}{2\overline{\alpha}+d}}, & \text{otherwise.} \end{cases}$$

Proof of Proposition 2.1.19. First, due to Lemma 2.1.18, we have for both cases (i) and (ii) for some constant $C \in \mathbb{R}_+$ and bandwidth $\boldsymbol{h} \in \mathbb{R}_+^d$ that

$$\|g_{\boldsymbol{h}}^{(p)} - g^{(p)}\|_{\mathbb{L}_d^2}^2 \leq C(h_1^{2\alpha_1} + \dots + h_d^{2\alpha_d}). \tag{2.23}$$

With Proposition 2.1.11, we obtain for some constant $C \in \mathbb{R}_+, \in \mathbb{R}_+^d$

$$\mathbb{E}\left[\|\widehat{g_{\boldsymbol{h}}^{(p)}} - g^{(p)}\|_{\mathbb{L}_d^2}^2\right] \leq C\left(h_1^{2\alpha_1} + \dots + h_d^{2\alpha_d} + \frac{1}{h_1 \cdots h_d n^p} + \frac{1}{n}\right).$$

Plugging the definition of $\boldsymbol{h}_{\text{opt}}$ into the first summand yields

$$\sum_{k \in [\![d]\!]} h_{k,\text{opt}}^{2\alpha_k} = \sum_{k \in [\![d]\!]} n^{-\frac{2\alpha_k \overline{\alpha} p}{\alpha_k(2\overline{\alpha}+d)}} = d n^{-\frac{2\overline{\alpha}p}{2\overline{\alpha}+d}}.$$

For the second summand we first note that

$$\begin{aligned} -p + \sum_{k \in [\![d]\!]} \frac{\overline{\alpha} p}{\alpha_k(2\overline{\alpha}+d)} &= -p + \frac{\overline{\alpha} p}{2\overline{\alpha}+d} \sum_{k \in [\![d]\!]} \frac{1}{\alpha_k} \\ &= \frac{-p(2\overline{\alpha}+d) + pd}{2\overline{\alpha}+d} = -\frac{2\overline{\alpha}p}{2\overline{\alpha}+d}. \end{aligned}$$

With this we obtain

$$\frac{1}{h_{1,\text{opt}} \cdots h_{d,\text{opt}} n^p} = n^{-p} \prod_{k \in [\![d]\!]} n^{\frac{\overline{\alpha} p}{\alpha_k(2\overline{\alpha}+d)}} = n^{-\frac{2\overline{\alpha}p}{2\overline{\alpha}+d}}.$$

Finally, this yields that there exists a constant $C \in \mathbb{R}_+$ such that

$$\begin{aligned} \mathbb{E}\left[\|\widehat{g_{\boldsymbol{h}_{\text{opt}}}^{(p)}} - g^{(p)}\|_{\mathbb{L}_d^2}^2\right] &\leq C \max\left(n^{-\frac{2\overline{\alpha}p}{2\overline{\alpha}+d}}, n^{-1}\right) \\ &= C \begin{cases} n^{-1}, & \text{if } \overline{\alpha} \geq \frac{d}{2(p-1)}, \\ n^{-\frac{2\overline{\alpha}p}{2\overline{\alpha}+d}}, & \text{otherwise.} \end{cases} \end{aligned}$$

This concludes the proof. □

Remark 2.1.20:
So far, we have considered the amount of components $p \in \mathbb{N}$ to be fixed. If $p_n \to \infty$ grows with the sample size n, it holds $\overline{\alpha} \geq \frac{d}{2(p-1)}$ for n large enough and the rate is of order n^{-1}. However, the factor $p^2 3^{2(p-2)} =: C_p$ is included in the constant of Proposition 2.1.11. Thus, $p_n \lesssim \log(n)$ needs to grow logarithmically in n or slower so that we still obtain $C_{p_n} n^{-1} \to 0$.
We have also assumed the dimension $d \in \mathbb{N}$ is fixed. Consider the case that $d_n \to \infty$ grows with the sample size and assume that the regularity is the same for all dimensions, i.e., $\alpha_k = \alpha$ for all $k \in [\![d_n]\!]$ and $n \in \mathbb{N}$. Then, $\overline{\alpha} = \alpha$ does not depend on d and it follows $\overline{\alpha} < \frac{d_n}{2(p-1)}$ for n large enough. Moreover, we obtain the rate $n^{-\frac{2\overline{\alpha}p}{2\overline{\alpha}p+d_n}}$, which gets slower for increasing d_n. Again d is included in the constant C. However, the dependence on d is not as obvious as the dependence on the parameter p.

Next, we deduce the regularity of $g^{(p)}$ from the regularities of the densities f_j, $j \in [\![p]\!]$ for both Sobolev and Nikolski spaces. That is, we show for both spaces, that if all densities f_j for $j \in [\![p]\!]$ belong to one regularity class, their convolution is also element of that regularity class. We first give the link ofr the Sobolev space.

Lemma 2.1.21:
If $f_j \in \mathcal{S}((\alpha_{j,1}, \dots, \alpha_{j,d}), R_j)$ for $j \in [\![p]\!]$, that is each component density belongs to a Sobolev space, then the convoluted density $g^{(p)}$ also belongs to a Sobolev space. More precisely, it holds $g^{(p)} \in \mathcal{S}((\sum_{j\in[\![p]\!]} \alpha_{j,1}, \dots, \sum_{j\in[\![p]\!]} \alpha_{j,d}), \prod_{j\in[\![p]\!]} R_j)$.

Proof of Lemma 2.1.21. The result follows analogously to the case $d = 1$ for $f_1 = \ldots = f_p$ derived in Chesneau et al. (2013). First note that it holds $g^{(p)} \in \mathbb{L}^2_d$ due to repetitively applying Young's inequality, see Property 1.1.3 (ii), and since $f_j \in \mathbb{L}^1_d \cap \mathbb{L}^2_d$ for all $j \in [\![p]\!]$. More precisely,

$$\|g^{(p)}\|_{\mathbb{L}^2_d} \leq \|f_p\|_{\mathbb{L}^2_d} \prod_{j\in[\![p-1]\!]} \|f_j\|_{\mathbb{L}^1_d} < \infty.$$

Further, note that, if $f_j \in \mathcal{S}((\alpha_{j,1}, \dots, \alpha_{j,d}), R_j)$ for $j \in [\![p]\!]$, it follows for any $\boldsymbol{u} \in \mathbb{R}^d$, $l \in [\![d]\!]$ and $j \in [\![p]\!]$ it holds

$$(1 + u_l^2)^{\alpha_{j,l}} |\mathcal{F}[f_j](\boldsymbol{u})|^2 \leq \max_{k\in[\![d]\!]} \sup_{\boldsymbol{t}\in\mathbb{R}^d} (1 + t_k^2)^{\alpha_{j,k}} |\mathcal{F}[f_j](\boldsymbol{t})|^2 \leq R_j^2.$$

This yields for any $\boldsymbol{t} \in \mathbb{R}^d$ and $k \in [\![d]\!]$ that

$$(1 + t_k^2)^{\sum_{j\in[\![p]\!]} \alpha_{j,k}} \prod_{j\in[\![p]\!]} |\mathcal{F}[f_j](\boldsymbol{t})|^2 \leq (1 + t_k^2)^{\alpha_{j,p}} |\mathcal{F}[f_p](\boldsymbol{t})|^2 \prod_{j\in[\![p-1]\!]} R_j^2.$$

With this, we get

$$
\begin{aligned}
\sum_{k\in[\![d]\!]} \int_{\mathbb{R}^d} &(1+t_k^2)^{\sum_{j\in[\![p]\!]}\alpha_{j,k}} |\mathcal{F}[g^{(p)}](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}) \\
&= \sum_{k\in[\![d]\!]} \int_{\mathbb{R}^d} (1+t_k^2)^{\sum_{j\in[\![p]\!]}\alpha_{j,k}} \prod_{j\in[\![p]\!]} |\mathcal{F}[f_j](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}) \\
&\le \left(\prod_{j\in[\![p-1]\!]} R_j\right) \sum_{k\in[\![d]\!]} \int_{\mathbb{R}^d} (1+t_k^2)^{\alpha_{p,k}} |\mathcal{F}[f_p](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}) \\
&\le \prod_{j\in[\![p]\!]} R_j^2.
\end{aligned}
$$

Finally, it holds that

$$
\max_{k\in[\![d]\!]} \sup_{\boldsymbol{t}\in\mathbb{R}^d} (1+t_k^2)^{\sum_{j\in[\![p]\!]}\alpha_{j,k}} |\mathcal{F}[g^{(p)}](\boldsymbol{t})|^2 \le \prod_{j\in[\![p]\!]} \max_{k\in[\![d]\!]} \sup_{\boldsymbol{t}\in\mathbb{R}^d} (1+t_k^2)^{\alpha_{j,k}} |\mathcal{F}[f_j](\boldsymbol{t})|^2 \le \prod_{j\in[\![p]\!]} R_j^2.
$$

Thus, $g^{(p)} \in \mathcal{S}((\sum_{j\in[\![p]\!]}\alpha_{j,1},\dots,\sum_{j\in[\![p]\!]}\alpha_{j,d}), \prod_{j\in[\![p]\!]} R_j)$, which completes the proof. □

Remark 2.1.22:
Consequently, in the case of anisotropic Sobolev spaces, the regularity parameter in the convergence rates is the harmonic mean of the directional regularities of the function $g^{(p)}$, i.e.

$$
\frac{1}{\overline{\alpha}} = \frac{1}{d} \sum_{k\in[\![d]\!]} \frac{1}{\sum_{j\in[\![p]\!]}\alpha_{j,k}}.
$$

Furthermore, if all directions and all components are of the same regularity $a \in \mathbb{R}_+$, that is $\alpha_{j,k} = a$ for all $j \in [\![p]\!], k \in [\![d]\!]$, the parameter simplifies to $\overline{\alpha} = pa$. In this case, the result of Proposition 2.1.19 yields: for $a \ge \frac{d}{2p(p-1)}$ we obtain the rate n^{-1}; otherwise the corresponding rate is of order $n^{-\frac{2ap^2}{2ap+d}}$.

Let us turn next to the link between the regularity of $g^{(p)}$ and its components for Nikolski classes.

Lemma 2.1.23:
If $f_j \in \mathcal{N}((\alpha_{j,1},\dots,\alpha_{j,d}), R_j)$ for $j \in [\![p]\!]$, that is, each component density belongs to a Nikolski class, and $\frac{\partial^{\lfloor\alpha_{j,k}\rfloor} f_j}{(\partial x_k)^{\lfloor\alpha_{j,k}\rfloor}}$ integrable for all $k \in [\![d]\!]$ and $j \in [\![p]\!]$, then the convoluted density $g^{(p)}$ also belongs to a Nikolski class. More precisely,there exists $R > 0$ such that $g^{(p)} \in \mathcal{N}(\boldsymbol{\delta}, R)$ with $\delta_k = \sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor + \max_{j\in[\![p]\!]}(\alpha_{j,k} - \lfloor\alpha_{j,k}\rfloor)$ for $k \in [\![d]\!]$.

Proof of Lemma 2.1.23. Fix $k \in [\![d]\!]$. First, note that for all $j \in [\![p]\!]$ the function f_j admits derivatives with respect to x_k up to order $\lfloor\alpha_{j,k}\rfloor$. Then,

$$
\frac{\partial^{\lfloor\alpha_{1,k}\rfloor} f_1}{(\partial x_k)^{\lfloor\alpha_{1,k}\rfloor}} * \dots * \frac{\partial^{\lfloor\alpha_{p,k}\rfloor} f_p}{(\partial x_k)^{\lfloor\alpha_{p,k}\rfloor}} = \frac{\partial^{\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor} g^{(p)}}{(\partial x_k)^{\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor}}.
$$

Thus, $g^{(p)}$ admits derivatives with respect to x_k up to order $\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor = \lfloor\delta_k\rfloor$. Define

$$R^2 := \max_{j\in[\![p]\!]}(R_j^2)\cdot \max_{k\in[\![d]\!]} \prod_{j\in[\![p]\!]}\left(\int_{\mathbb{R}^d}\left|\frac{\partial^{\lfloor\alpha_{j,k}\rfloor}f_j}{(\partial x_k)^{\lfloor\alpha_{j,k}\rfloor}}(\boldsymbol{x})\right| d\lambda^d(\boldsymbol{x})\vee 1\right)^2.$$

Note that by Young's inequality, see Property 1.1.3 (ii), it holds

$$\begin{aligned}
\|\frac{\partial^{\sum_{j\in[\![p]\!]},\lfloor\alpha_{j,k}\rfloor}g^{(p)}}{(\partial x_k)^{\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor}}\|_{\mathbb{L}_d^2}^2 &\le \|\frac{\partial^{\sum_{j=2}^p\lfloor\alpha_{j,k}\rfloor}f_2*\cdots*f_p}{(\partial x_k)^{\sum_{j=2}^p\lfloor\alpha_{j,k}\rfloor}}\|_{\mathbb{L}_d^1}^2\|\frac{\partial^{\lfloor\alpha_{1,k}\rfloor}f_1}{(\partial x_k)^{\lfloor\alpha_{1,k}\rfloor}}\|_{\mathbb{L}_d^2}^2\\
&\le R_1^2\prod_{j=2}^p\|\frac{\partial^{\lfloor\alpha_{j,k}\rfloor}f_j}{(\partial x_k)^{\lfloor\alpha_{j,k}\rfloor}}\|_{\mathbb{L}_d^1}^2 \le R^2.
\end{aligned}$$

Assume without loss of generality that $\max_{j\in[\![p]\!]}(\alpha_{j,k}-\lfloor\alpha_{j,k}\rfloor) = \alpha_{1,k}-\lfloor\alpha_{1,k}\rfloor$. Then, it holds that $\delta_k - \lfloor\delta_k\rfloor = \alpha_{1,k}-\lfloor\alpha_{1,k}\rfloor$. Setting for $y\in\mathbb{R}$

$$f_{1,y}(\boldsymbol{x}) := \frac{\partial^{\lfloor\alpha_{1,k}\rfloor}f_1}{(\partial x_k)^{\lfloor\alpha_{1,k}\rfloor}}(x_1,\dots,x_{k-1},x_k+y,x_{k+1},\dots,x_d),$$

splitting the convolution, it follows that

$$\begin{aligned}
&\int_{\mathbb{R}^d}\left(\frac{\partial^{\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor}g^{(p)}}{(\partial x_k)^{\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor}}(x_1,\dots,x_{j-1},x_j+y,x_{j+1},\dots,x_d) - \frac{\partial^{\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor}g^{(p)}}{(\partial x_k)^{\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor}}(\boldsymbol{x})\right)^2 d\lambda^d(\boldsymbol{x})\\
&= \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}(\frac{\partial^{\lfloor\alpha_{1,k}\rfloor}f_1}{(\partial x_k)^{\lfloor\alpha_{1,k}\rfloor}}((x_1,\dots,x_{j-1},x_k+y,x_{k+1},\dots,x_d)-\boldsymbol{u}) - \frac{\partial^{\lfloor\alpha_{1,k}\rfloor}f_1}{(\partial x_k)^{\lfloor\alpha_{1,k}\rfloor}}(\boldsymbol{x}-\boldsymbol{u}))\right.\\
&\qquad\left.\cdot\frac{\partial^{\sum_{j=2}^p\lfloor\alpha_{j,k}\rfloor}(f_2*\cdots*f_p)}{(\partial x_k)^{\sum_{j=2}^p\lfloor\alpha_{j,k}\rfloor}}(\boldsymbol{u})d\lambda^d(\boldsymbol{u})\right)^2 d\lambda^d(\boldsymbol{x})\\
&= \|(f_{1,y}-\frac{\partial^{\lfloor\alpha_{1,k}\rfloor}f_1}{(\partial x_k)^{\lfloor\alpha_{1,k}\rfloor}})*\frac{\partial^{\sum_{j=2}^p\lfloor\alpha_{j,k}\rfloor}(f_2*\cdots*f_p)}{(\partial x_k)^{\sum_{j=2}^p\lfloor\alpha_{j,k}\rfloor}}\|_{\mathbb{L}_d^2}^2.
\end{aligned}$$

Applying again Young's inequality, see Property 1.1.3 (ii), it holds

$$\begin{aligned}
&\|(f_{1,y}-\frac{\partial^{\lfloor\alpha_{1,k}\rfloor}f_1}{(\partial x_k)^{\lfloor\alpha_{1,k}\rfloor}})*\frac{\partial^{\sum_{j=2}^p\lfloor\alpha_{j,k}\rfloor}(f_2*\cdots*f_p)}{(\partial x_k)^{\sum_{j=2}^p\lfloor\alpha_{j,k}\rfloor}}\|_{\mathbb{L}_d^2}^2\\
&\le \|\frac{\partial^{\sum_{j=2}^p\lfloor\alpha_{j,k}\rfloor}(f_2*\cdots*f_p)}{(\partial x_k)^{\sum_{j=2}^p\lfloor\alpha_{j,k}\rfloor}}\|_{\mathbb{L}_d^1}^2\|f_{1,y}-\frac{\partial^{\lfloor\alpha_{1,k}\rfloor}f_1}{(\partial x_k)^{\lfloor\alpha_{1,k}\rfloor}}\|_{\mathbb{L}_d^2}^2,
\end{aligned}$$

and using $f_j\in\mathcal{N}((\alpha_{j,1},\dots,\alpha_{j,d}),R_j)$ for $j\in[\![p]\!]$ we conclude

$$\begin{aligned}
&\|(f_{1,y}-\frac{\partial^{\lfloor\alpha_{1,k}\rfloor}f_1}{(\partial x_k)^{\lfloor\alpha_{1,k}\rfloor}})*\frac{\partial^{\sum_{j=2}^p\lfloor\alpha_{j,k}\rfloor}(f_2*\cdots*f_p)}{(\partial x_k)^{\sum_{j=2}^p\lfloor\alpha_{j,k}\rfloor}}\|_{\mathbb{L}_d^2}^2\\
&\le \prod_{j=2}^p\|\frac{\partial^{\lfloor\alpha_{j,k}\rfloor}f_j}{(\partial x_k)^{\lfloor\alpha_{j,k}\rfloor}}\|_{\mathbb{L}_d^1}^2 R_1^2|y|^{2(\alpha_{1,k}-\lfloor\alpha_{1,k}\rfloor)} \le R|y|^{2(\delta_k-\lfloor\delta_k\rfloor)}.
\end{aligned}$$

Consequently, $g^{(p)}\in\mathcal{N}(\boldsymbol{\delta},R)$. This completes the proof. □

2.1.4 Point-wise quadratic risk

In this section, we discuss the point-wise quadratic risk generalizing Proposition 2.1.6 and derive convergence rates for Hölder classes.

Analogously to the case of $p, d = 1$ discussed in Section 2.1.1, note that if a density f is bounded, it is also in $\mathbb{L}^2_d$ since

$$\int_{\mathbb{R}^d} |f(\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}) \leq \|f\|_{\mathbb{L}^\infty_d} \int_{\mathbb{R}} f(\boldsymbol{t}) d\lambda^d(\boldsymbol{t}) = \|f\|_{\mathbb{L}^\infty_d} < \infty.$$

Consequently, if $f_1, \dots, f_p \in \mathbb{L}^\infty_d$ it follows that $f_1, \dots, f_p \in \mathbb{L}^2_d$ and, thus, $g^{(p)} \in \mathbb{L}^2_d$. Additionally, it follows that the Fourier transform of their convolution $\mathcal{F}[g^{(p)}]$ is in $\mathbb{L}^1_d$. More precisely, since $|\mathcal{F}[f_j](x)| \leq 1$ for all $x \in \mathbb{R}^d$ and $j \in [\![p]\!]$, see Property 1.1.3 (iii), it holds that

$$\begin{aligned}
\int_{\mathbb{R}^d} |\mathcal{F}[g^{(p)}](\boldsymbol{x})| d\lambda^d(\boldsymbol{x}) &= \int_{\mathbb{R}^d} \left| \prod_{j \in [\![p]\!]} \mathcal{F}[f_j](\boldsymbol{x}) \right| d\lambda^d(\boldsymbol{x}) \\
&\leq \left(\int_{\mathbb{R}^d} |\mathcal{F}[f_1](\boldsymbol{x})|^2 d\lambda^d(\boldsymbol{x}) \right)^{1/2} \left(\int_{\mathbb{R}^d} |\mathcal{F}[f_2](\boldsymbol{x})|^2 d\lambda^d(\boldsymbol{x}) \right)^{1/2} \\
&= \|f_1\|_{\mathbb{L}^2_d} \|f_2\|_{\mathbb{L}^2_d} < \infty.
\end{aligned}$$

Finally, it also follows that $g^{(p)} \in \mathbb{L}^\infty_d$ since for any $\boldsymbol{x} \in \mathbb{R}^d$ we have

$$\begin{aligned}
|g^{(p)}(\boldsymbol{x})| = |\mathcal{F}^\dagger[\mathcal{F}[g^{(p)}]](\boldsymbol{x})| &= \left| \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle \boldsymbol{x}, \boldsymbol{t} \rangle} \mathcal{F}[g^{(p)}](\boldsymbol{t}) d\lambda^d(\boldsymbol{t}) \right| \\
&\leq \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\mathcal{F}[g^{(p)}](\boldsymbol{t})| d\lambda^d(\boldsymbol{t}) = \frac{1}{(2\pi)^d} \|\mathcal{F}[g^{(p)}]\|_{\mathbb{L}^1_d} < \infty.
\end{aligned}$$

Consequently, point evaluation of $g^{(p)}$ is well-defined and we also provide a point-wise risk bound.

Proposition 2.1.24 (Bound of the point-wise quadratic risk):
Let Assumption 2.1.10 be satisfied and $p \geq 3$. If $f_j \in \mathbb{L}^\infty_d$ for all $j \in [\![p]\!]$, then, for $\boldsymbol{x} \in \mathbb{R}^d$ the estimator $\widehat{g^{(p)}_{\boldsymbol{h}}}(\boldsymbol{x})$ of $g^{(p)}(\boldsymbol{x})$ defined in Equation (2.11) satisfies that there exist $C_1, C_2 > 0$ such that

$$\mathbb{E}\left[(\widehat{g^{(p)}_{\boldsymbol{h}}}(\boldsymbol{x}) - g^{(p)}(\boldsymbol{x}))^2 \right] \leq (g^{(p)}_{\boldsymbol{h}}(\boldsymbol{x}) - g^{(p)}(\boldsymbol{x}))^2 + C_1 \frac{1}{h_1 \cdots h_d n_1 \cdots n_p} + C_2 \sum_{j \in [\![p]\!]} \frac{1}{n_j}.$$

Remark 2.1.25:
For the case $p = 2$, analogously to the proof of Proposition 2.1.6, it can be shown that: If it holds $f_1, f_2 \in \mathbb{L}^\infty_d$ and $L \in \mathbb{L}^1_d \cap \mathbb{L}^2_d$ is an even function, then, the estimator $\widehat{g^{(2)}_{\boldsymbol{h}}}$ of $g^{(2)}$ defined by (2.11) satisfies for $\boldsymbol{x} \in \mathbb{R}^d$ that

$$\mathbb{E}\left[(\widehat{g^{(2)}_{\boldsymbol{h}}}(\boldsymbol{x}) - g^{(2)}(\boldsymbol{x}))^2 \right] \leq (g^{(2)}_{\boldsymbol{h}}(\boldsymbol{x}) - g^{(2)}(\boldsymbol{x}))^2 + \|L\|_{\mathbb{L}^2_d} \frac{\|g^{(2)}\|_{\mathbb{L}^\infty_d}}{n_1 n_2 h_1 h_2} + C_f \|L\|^2_{\mathbb{L}^1} \left(\frac{1}{n_1} + \frac{1}{n_2} \right),$$

where $C_f := \|f_1\|_{\mathbb{L}^\infty_d} \|f_2\|^2_{\mathbb{L}^2_d} + \|f_2\|_{\mathbb{L}^\infty_d} \|f_1\|^2_{\mathbb{L}^2_d}$.

Proof of Proposition 2.1.24. For all $\boldsymbol{x} \in \mathbb{R}^d$ we have that $\mathbb{E}\left[\widehat{g_{\boldsymbol{h}}^{(p)}}(\boldsymbol{x})\right] = g_{\boldsymbol{h}}^{(p)}(\boldsymbol{x})$, see Equation (2.16). Thus, we get the standard decomposition for $\boldsymbol{x} \in \mathbb{R}^d$

$$\mathbb{E}\left[(\widehat{g_{\boldsymbol{h}}^{(p)}}(\boldsymbol{x}) - g^{(p)}(\boldsymbol{x}))^2\right] = (g_{\boldsymbol{h}}^{(p)}(\boldsymbol{x}) - g^{(p)}(\boldsymbol{x}))^2 + \mathbb{E}\left[(\widehat{g_{\boldsymbol{h}}^{(p)}}(\boldsymbol{x}) - g_{\boldsymbol{h}}^{(p)}(\boldsymbol{x}))^2\right].$$

For the variance we get

$$\begin{aligned}
&\mathbb{E}\left[(\widehat{g_{\boldsymbol{h}}^{(p)}}(\boldsymbol{x}) - g_{\boldsymbol{h}}^{(p)}(\boldsymbol{x}))^2\right] \\
&= \mathbb{E}\left[\left(\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-i\langle \boldsymbol{t},\boldsymbol{x}\rangle}\mathcal{F}[L](\boldsymbol{h}\boldsymbol{t})\left(\prod_{j\in[\![p]\!]}\frac{1}{n_j}\sum_{k=1}^{n_j} e^{i\langle \boldsymbol{t},\boldsymbol{X}_k^{(j)}\rangle} - \prod_{j\in[\![p]\!]}\mathcal{F}[f_j](\boldsymbol{t})\right) d\lambda^d(\boldsymbol{t})\right)^2\right] \\
&\leq \frac{1}{(2\pi)^{2d}}\prod_{j\in[\![p]\!]}\frac{1}{n_j^2}\sum_{\substack{k_1\in[\![n_1]\!],\ldots,\\ k_p\in[\![n_p]\!]}}\sum_{\substack{l_1\in[\![n_1]\!],\ldots,\\ l_p\in[\![n_p]\!]}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|\mathcal{F}[L](\boldsymbol{h}\boldsymbol{t})\mathcal{F}[L](\boldsymbol{h}\boldsymbol{s})| \\
&\qquad \left|\mathbb{E}\left[\left(e^{i\langle \boldsymbol{t},\boldsymbol{X}_{k_1}^{(1)}+\ldots+\boldsymbol{X}_{k_p}^{(p)}\rangle} - \prod_{j\in[\![p]\!]}\mathcal{F}[f_j](\boldsymbol{t})\right)\left(e^{i\langle \boldsymbol{s},\boldsymbol{X}_{l_1}^{(1)}+\ldots+\boldsymbol{X}_{l_p}^{(p)}\rangle} - \prod_{j\in[\![p]\!]}\mathcal{F}[f_j](\boldsymbol{s})\right)\right]\right| \\
&\qquad d\lambda^d(\boldsymbol{t})\, d\lambda^d(\boldsymbol{s}).
\end{aligned} \tag{2.24}$$

The expectation is equal to zero if $k_j \neq l_j$ for all $j \in [\![p]\!]$. Consequently, we consider the case that for some $\tilde{p} \in [\![p]\!]$ it holds that $\tilde{p}$-many index-couples are the same. For this, we start by considering the case that the first indices are the same, i.e.,

$$\begin{aligned}
k_1 = l_1, \ldots, k_{\tilde{p}} = l_{\tilde{p}}, \\
k_{\tilde{p}+1} \neq l_{\tilde{p}+1}, \ldots, k_p \neq l_p.
\end{aligned}$$

Before considering Equation (2.24), let us first note that due to independence between and within samples and using that the first $\tilde{p}$ indices are the same it follows for $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^d$ that

$$\begin{aligned}
&\mathbb{E}\left[e^{i\langle \boldsymbol{t},\boldsymbol{X}_{k_1}^{(1)}+\ldots+\boldsymbol{X}_{k_p}^{(p)}\rangle} e^{i\langle \boldsymbol{s},\boldsymbol{X}_{l_1}^{(1)}+\ldots+\boldsymbol{X}_{l_p}^{(p)}\rangle}\right] \\
&= \mathbb{E}\left[e^{i\langle \boldsymbol{t}+\boldsymbol{s},\boldsymbol{X}_{k_1}^{(1)}+\ldots+\boldsymbol{X}_{k_{\tilde{p}}}^{(\tilde{p})}\rangle}\right]\mathbb{E}\left[e^{i\langle \boldsymbol{t},\boldsymbol{X}_{k_{\tilde{p}+1}}^{(\tilde{p}+1)}+\ldots+\boldsymbol{X}_{k_p}^{(p)}\rangle}\right]\mathbb{E}\left[e^{i\langle \boldsymbol{s},\boldsymbol{X}_{l_{\tilde{p}+1}}^{(\tilde{p}+1)}+\ldots+\boldsymbol{X}_{l_p}^{(p)}\rangle}\right] \\
&= \prod_{j\in[\![\tilde{p}]\!]}\mathbb{E}\left[e^{i\langle \boldsymbol{t}+\boldsymbol{s},\boldsymbol{X}_{k_j}^{(j)}\rangle}\right]\prod_{j\in[\![\tilde{p}+1]\!]}\mathbb{E}\left[e^{i\langle \boldsymbol{t},\boldsymbol{X}_{k_j}^{(j)}\rangle}\right]\mathbb{E}\left[e^{i\langle \boldsymbol{s},\boldsymbol{X}_{l_j}^{(j)}\rangle}\right] \\
&= \prod_{j\in[\![\tilde{p}]\!]}\mathcal{F}[f_j](\boldsymbol{s}+\boldsymbol{t})\prod_{j=\tilde{p}+1}^{p}\mathcal{F}[f_j](\boldsymbol{s})\mathcal{F}[f_j](\boldsymbol{t}).
\end{aligned}$$

Further, we have that

$$\mathbb{E}\left[\left(e^{i\langle \boldsymbol{t}, \boldsymbol{X}_{k_1}^{(1)}+\ldots+\boldsymbol{X}_{k_p}^{(p)}\rangle} - \prod_{j\in[\![p]\!]} \mathcal{F}[f_j](\boldsymbol{t})\right)\left(e^{i\langle \boldsymbol{s}, \boldsymbol{X}_{l_1}^{(1)}+\ldots+\boldsymbol{X}_{l_p}^{(p)}\rangle} - \prod_{j\in[\![p]\!]} \mathcal{F}[f_j](\boldsymbol{s})\right)\right]$$
$$= \mathbb{E}\left[e^{i\langle \boldsymbol{t}, \boldsymbol{X}_{k_1}^{(1)}+\ldots+\boldsymbol{X}_{k_p}^{(p)}\rangle} e^{i\langle \boldsymbol{s}, \boldsymbol{X}_{l_1}^{(1)}+\ldots+\boldsymbol{X}_{l_p}^{(p)}\rangle}\right] - \prod_{j\in[\![p]\!]} \mathcal{F}[f_j](\boldsymbol{s})\mathcal{F}[f_j](\boldsymbol{t})$$
$$= \prod_{j\in[\![\tilde{p}]\!]} \mathcal{F}[f_j](\boldsymbol{s}+\boldsymbol{t}) \prod_{j=\tilde{p}+1}^{p} \mathcal{F}[f_j](\boldsymbol{s})\mathcal{F}[f_j](\boldsymbol{t}) - \prod_{j\in[\![p]\!]} \mathcal{F}[f_j](\boldsymbol{s})\mathcal{F}[f_j](\boldsymbol{t}).$$

Inserting this calculation into Equation (2.24), it follows that

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |\mathcal{F}[L](\boldsymbol{ht})\mathcal{F}[L](\boldsymbol{hs})|$$
$$\left|\mathbb{E}\left[\left(e^{i\langle \boldsymbol{t}, \boldsymbol{X}_{k_1}^{(1)}+\ldots+\boldsymbol{X}_{k_p}^{(p)}\rangle} - \prod_{j\in[\![p]\!]} \mathcal{F}[f_j](\boldsymbol{t})\right)\left(e^{i\langle \boldsymbol{s}, \boldsymbol{X}_{l_1}^{(1)}+\ldots+\boldsymbol{X}_{l_p}^{(p)}\rangle} - \prod_{j\in[\![p]\!]} \mathcal{F}[f_j](\boldsymbol{s})\right)\right]\right| d\lambda^d(\boldsymbol{t})\, d\lambda^d(\boldsymbol{s})$$
$$\leq \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |\mathcal{F}[L](\boldsymbol{ht})\mathcal{F}[L](\boldsymbol{hs})| \left|\prod_{j\in[\![\tilde{p}]\!]} \mathcal{F}[f_j](\boldsymbol{s}+\boldsymbol{t}) \prod_{j=\tilde{p}+1}^{p} \mathcal{F}[f_j](\boldsymbol{s})\mathcal{F}[f_j](\boldsymbol{t})\right| d\lambda^d(\boldsymbol{t})\, d\lambda^d(\boldsymbol{s})$$
$$+ \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |\mathcal{F}[L](\boldsymbol{ht})\mathcal{F}[L](\boldsymbol{hs})| \left|\prod_{j\in[\![p]\!]} \mathcal{F}[f_j](\boldsymbol{s})\mathcal{F}[f_j](\boldsymbol{t})\right| d\lambda^d(\boldsymbol{t})\, d\lambda^d(\boldsymbol{s}) =: T_1 + T_2.$$

For the second summand T_2, with Cauchy–Schwarz inequality and boundedness of the Fourier transform, see Property 1.1.3 (iii), $|\mathcal{F}[f_j](\boldsymbol{t})| \leq 1$ for all $j \in [\![p]\!]$ and $\boldsymbol{t} \in \mathbb{R}^d$ we get that

$$T_2 \leq \|\mathcal{F}[L]\|^2_{\mathbb{L}^\infty_d} \left|\int_{\mathbb{R}^d} |\mathcal{F}[f_1](\boldsymbol{s})\mathcal{F}[f_2](\boldsymbol{s})|\, d\lambda^d(\boldsymbol{s})\right|^2$$
$$\leq \|\mathcal{F}[L]\|^2_{\mathbb{L}^\infty_d} \max_{j,k\in[\![p]\!], j\neq k} \|\mathcal{F}[f_j]\|^2_{\mathbb{L}^2_d} \|\mathcal{F}[f_k]\|^2_{\mathbb{L}^2_d} =: \tilde{C}_0.$$

For T_1, we consider different cases depending on $\tilde{p}$, recalling that we assume $p \geq 3$.

Case $\tilde{p} \in [\![p-2]\!]$. In this case we apply again Cauchy–Schwarz inequality and get

$$T_1 \leq \|\mathcal{F}[L]\|^2_{\mathbb{L}^\infty_d} \left|\int_{\mathbb{R}^d} |\mathcal{F}[f_{p-1}](\boldsymbol{t})\mathcal{F}[f_p](\boldsymbol{t})| d\lambda^d(\boldsymbol{t})\right|^2$$
$$\leq \|\mathcal{F}[L]\|^2_{\mathbb{L}^\infty_d} \max_{j,k\in[\![p]\!], j\neq k} \|\mathcal{F}[f_j]\|^2_{\mathbb{L}^2_d} \|\mathcal{F}[f_k]\|^2_{\mathbb{L}^2_d} =: \tilde{C}_1.$$

Case $\tilde{p} = p - 1$. Applying Lemma 2.4.3 given in Section 2.4 with $h_1(\boldsymbol{t}) = \mathcal{F}[L](\boldsymbol{ht})\mathcal{F}[f_p](\boldsymbol{t})$ and $h_2(\boldsymbol{t}) = \prod_{j\in[\![p-1]\!]} \mathcal{F}[f_j](\boldsymbol{t})$ for $\boldsymbol{t} \in \mathbb{R}^d$ and, subsequently, Cauchy–Schwarz inequality for the first summand it follows

$$
\begin{aligned}
T_1 &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |\mathcal{F}[L](\boldsymbol{ht})\mathcal{F}[f_p](\boldsymbol{t})\mathcal{F}[L](\boldsymbol{hs})\mathcal{F}[f_p](\boldsymbol{s})| \left|\prod_{j\in[\![p-1]\!]} \mathcal{F}[f_j](\boldsymbol{s}+\boldsymbol{t})\right| d\lambda^d(\boldsymbol{t})\, d\lambda^d(\boldsymbol{s}) \\
&\le \int_{\mathbb{R}^d} |\mathcal{F}[L](\boldsymbol{ht})\mathcal{F}[f_p](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}) \int_{\mathbb{R}^d} \left|\prod_{j\in[\![p-1]\!]} \mathcal{F}[f_j](\boldsymbol{s})\right| d\lambda^d(\boldsymbol{s}) \\
&\le \|\mathcal{F}[L]\|^2_{\mathbb{L}^\infty_d} \|\mathcal{F}[f_p]\|_{\mathbb{L}^2_d} \|\mathcal{F}[f_1]\|_{\mathbb{L}^2_d} \|\mathcal{F}[f_2]\|_{\mathbb{L}^2_d} \\
&\le \|\mathcal{F}[L]\|^2_{\mathbb{L}^\infty_d} \max_{i,j,k\in[\![p]\!], i\neq j\neq k} \|\mathcal{F}[f_i]\|_{\mathbb{L}^2_d} \|\mathcal{F}[f_j]\|_{\mathbb{L}^2_d} \|\mathcal{F}[f_k]\|_{\mathbb{L}^2_d} =: \tilde{C}_2.
\end{aligned}
$$

Case $\tilde{p} = p$. In this case, with change of variables and again Cauchy–Schwarz inequality we get

$$
\begin{aligned}
T_1 &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |\mathcal{F}[L](\boldsymbol{ht})\mathcal{F}[L](\boldsymbol{hs})| \left|\prod_{j\in[\![p]\!]} \mathcal{F}[f_j](\boldsymbol{s}+\boldsymbol{t})\right| d\lambda^d(\boldsymbol{t})\, d\lambda^d(\boldsymbol{s}) \\
&\le \|\mathcal{F}[L]\|_{\mathbb{L}^\infty_d} \int_{\mathbb{R}^d} |\mathcal{F}[L](\boldsymbol{ht})| \int_{\mathbb{R}^d} |\mathcal{F}[f_1](\boldsymbol{s})\mathcal{F}[f_2](\boldsymbol{s})| d\lambda^d(\boldsymbol{s})\, d\lambda^d(\boldsymbol{t}) \\
&\le \|\mathcal{F}[L]\|_{\mathbb{L}^\infty_d} \frac{1}{h_1\cdots h_d} \|\mathcal{F}[L]\|_{\mathbb{L}^1_d} \max_{j,k\in[\![p]\!], j\neq k} \|\mathcal{F}[f_j]\|^2_{\mathbb{L}^2_d} \|\mathcal{F}[f_k]\|^2_{\mathbb{L}^2_d} \\
&=: \frac{1}{h_1\cdots h_d}\tilde{C}_3.
\end{aligned}
$$

Recall that for simplicity, we assumed that the first indices are the same. Due to symmetry, similar calculations hold for any $\tilde{p}$-many index-couples being the same, adjusting the index sets accordingly. In addition, since the bounds do not depend on $\tilde{p}$ or the ordering of the indices, we use the bound from one of the cases for each summand in Equation (2.24). It remains to determine the amount of summands that appear for each case.

The amount of summands Equation (2.24) corresponding to Case 3, i.e., with $k_j = l_j$ for all $j \in [\![p]\!]$ is equal to $\prod_{j\in[\![p]\!]} n_j$. Further, the amount of summands corresponding either to Case 1 or 2 can be upper bounded by $(p-1)p!(\prod_{j\in[\![p]\!]} n_j^2) \sum_{j\in[\![p]\!]} \frac{1}{n_j}$, see Lemma 2.4.4.

Plugging all calculations into Equation (2.24) we finally get

$$
\begin{aligned}
&\mathbb{E}\left[(\widehat{g^{(p)}_{\boldsymbol{h}}}(\boldsymbol{x}) - g^{(p)}_{\boldsymbol{h}}(\boldsymbol{x}))^2\right] \\
&\le \frac{1}{(2\pi)^{2d}} \left(\prod_{j\in[\![p]\!]} \frac{1}{n_j^2}\right) \left((\tilde{C}_0 + \frac{1}{\boldsymbol{h}}\tilde{C}_3) \prod_{j\in[\![p]\!]} n_j + (\tilde{C}_0 + \tilde{C}_1 + \tilde{C}_2)(p-1)p!(\prod_{j\in[\![p]\!]} n_j^2) \sum_{j\in[\![p]\!]} \frac{1}{n_j} \right) \\
&= \frac{\tilde{C}_3}{(2\pi)^{2d}} \frac{1}{h_1\cdots h_d n_1\cdots n_p} + \frac{\tilde{C}_0}{(2\pi)^{2d}} \prod_{j\in[\![p]\!]} \frac{1}{n_j} + \frac{(\tilde{C}_1 + \tilde{C}_2)(p-1)p!}{(2\pi)^{2d}} \sum_{j\in[\![p]\!]} \frac{1}{n_j} \\
&\le C_1 \frac{1}{h_1\cdots h_d n_1\cdots n_p} + C_2 \sum_{j\in[\![p]\!]} \frac{1}{n_j},
\end{aligned}
$$

with $C_1 := \frac{\tilde{C}_3}{(2\pi)^{2d}}$ and $C_2 := \frac{\tilde{C}_0 + (\tilde{C}_1+\tilde{C}_2)(p-1)p!}{(2\pi)^{2d}}$. This completes the result. □

Convergence rates

Let us derive convergence rates for the kernel estimator defined in Equation (2.13) based on its point-wise quadratic risk. More precisely, we discuss the order of the upper bound given in Proposition 2.1.24. For this, we consider the case of Hölder classes. We state the results by imposing the regularity on $g^{(p)}$ and give a link to the regularity of the component densities at the end of this section. First, we give a definition of Hölder classes for d dimensions.

Definition 2.1.26 (Hölder class):
A function $h\colon \mathbb{R}^d \to \mathbb{R}$ belongs to Hölder class $\Sigma(\boldsymbol{\alpha}, R)$ for $R \in \mathbb{R}_+$ and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d) \in \mathbb{R}_+^d$ if f admits derivatives with respect to x_k up to order $\lfloor \alpha_k \rfloor$ for $k \in [\![d]\!]$ and if for any $\boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ and $x_k' \in \mathbb{R}$ it holds that

$$\left| \frac{\partial^{\lfloor \alpha_k \rfloor} h}{(\partial x_k)^{\lfloor \alpha_k \rfloor}}(x_1, \dots, x_{k-1}, x_k', x_{k+1}, \dots, x_d) - \frac{\partial^{\lfloor \alpha_k \rfloor} h}{(\partial x_k)^{\lfloor \alpha_k \rfloor}}(\boldsymbol{x}) \right| \leq R|x_k' - x_k|^{\alpha_k - \lfloor \alpha_k \rfloor}.$$

Analogously to the integrated quadratic risk, we show for the squared bias the following result. We need the same assumptions on L as for Nikolski classes.

Lemma 2.1.27:
Let Assumption 2.1.10 and Assumption 2.1.17 be satisfied, $\boldsymbol{\alpha} \in \mathbb{R}_+^d$ and $R \in \mathbb{R}_+$. If L is of order $(\lfloor \alpha_1 \rfloor, \dots, \lfloor \alpha_d \rfloor)$, $g^{(p)} \in \Sigma(\boldsymbol{\alpha}, R)$ and , then, it holds for the bias that there exists a constant $C \in \mathbb{R}_+$ such that for $\boldsymbol{x} \in \mathbb{R}^d$

$$(g_{\boldsymbol{h}}^{(p)}(\boldsymbol{x}) - g^{(p)}(\boldsymbol{x}))^2 \leq C(h_1^{2\alpha_1} + \ldots + h_d^{2\alpha_d}).$$

Proof of Lemma 2.1.27. The following computations are standard, see for example Tsybakov (2008), and we give the proofs for the sake of completeness. Using Equation (2.21) we have that for $\boldsymbol{x} \in \mathbb{R}^d$

$$g_{\boldsymbol{h}}^{(p)}(\boldsymbol{x}) - g^{(p)}(\boldsymbol{x}) = \int_{\mathbb{R}^d} L(\boldsymbol{u})(g^{(p)}(\boldsymbol{h}\boldsymbol{u} + \boldsymbol{x}) - g^{(p)}(\boldsymbol{x}))d\lambda^d(\boldsymbol{u}).$$

Analogously to the proof of Lemma 2.1.18, the idea is to apply component-wise Taylor's Lagrange formula on each summand. More precisely, for $\boldsymbol{u} \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&g_{\boldsymbol{h}}^{(p)}(\boldsymbol{u}\boldsymbol{h} + \boldsymbol{x}) - g^{(p)}(\boldsymbol{x}) \\
&= \sum_{j \in [\![d]\!]} \left(\sum_{k=1}^{\lfloor \alpha_j \rfloor - 1} \frac{(u_j h_j)^k}{k!} \frac{\partial^k g^{(p)}}{(\partial x_j)^k}(x_1, \dots, x_j, u_{j+1}h_{j+1} + x_{j+1}, \dots, u_d h_d + x_d) \right. \\
&\qquad \left. + \frac{(u_j h_j)^{\lfloor \alpha_j \rfloor}}{\lfloor \alpha_j \rfloor !} \frac{\partial^{\lfloor \alpha_j \rfloor} g^{(p)}}{(\partial x_j)^{\lfloor \alpha_j \rfloor}}(x_1, \dots, x_{j-1}, u_j h_j + x_j, \dots, u_d h_d + x_d) \right).
\end{aligned}$$

Since L is an even function of order $(\lfloor \alpha_1 \rfloor, \dots, \lfloor \alpha_d \rfloor)$ and $g^{(p)} \in \Sigma(\boldsymbol{\alpha}, R)$ satisfying Assumption 2.1.17 it follows

$$\begin{aligned}
&\int_{\mathbb{R}^d} L(\boldsymbol{u})(g_{\boldsymbol{h}}^{(p)}(\boldsymbol{u}\boldsymbol{h}+\boldsymbol{x})-g^{(p)}(\boldsymbol{x}))d\lambda^d(\boldsymbol{u}) \\
&= \sum_{j\in[\![d]\!]} \int_{\mathbb{R}^d} L(\boldsymbol{u}) \frac{(u_j h_j)^{\lfloor \alpha_j \rfloor}}{\lfloor \alpha_j \rfloor !} \Bigg(\frac{\partial^{\lfloor \alpha_j \rfloor} g^{(p)}}{(\partial x_j)^{\lfloor \alpha_j \rfloor}}(x_1,\dots,x_{j-1},u_j h_j + x_j,\dots,u_d h_d + x_d) \\
&\qquad\qquad - \frac{\partial^{\lfloor \alpha_j \rfloor} g^{(p)}}{(\partial x_j)^{\lfloor \alpha_j \rfloor}}(x_1,\dots,x_j,u_{j+1} h_{j+1} + x_{j+1},\dots,u_d h_d + x_d) \Bigg) d\lambda^d(\boldsymbol{u}) \\
&\le \sum_{j\in[\![d]\!]} \frac{R h_j^{\alpha_j}}{\lfloor \alpha_j \rfloor !} \int_{\mathbb{R}^d} |L(\boldsymbol{u}) u_j^{\alpha_j}| d\lambda^d(\boldsymbol{u}) \le \sum_{j\in[\![d]\!]} \frac{R}{\lfloor \alpha_j \rfloor !} \int_{\mathbb{R}^d} |L(\boldsymbol{u}) u_j^{\alpha_j}| d\lambda^d(\boldsymbol{u}) \sum_{j\in[\![d]\!]} h_j^{\alpha_j}.
\end{aligned}$$

With the last inequality we get the result, i.e.,

$$(g_{\boldsymbol{h}}^{(p)}(\boldsymbol{x})-g^{(p)}(\boldsymbol{x}))^2 \le 2C \sum_{j\in[\![d]\!]} h_j^{2\alpha_j},$$

with $C := \sum_{j\in[\![d]\!]} \frac{R}{\lfloor \alpha_j \rfloor !} \int_{\mathbb{R}^d} |L(\boldsymbol{u}) u_j^{\alpha_j}| d\lambda^d(\boldsymbol{u})$. □

Analogously to the integrated quadratic risk, the following result for the rate of convergence can be shown. We omit the proof here.

Proposition 2.1.28:
Let Assumption 2.1.10 and Assumption 2.1.17 be satisfied. Assume that all n_j are of the same order $n \in \mathbb{N}$, that is, $n_j = c_j n$ for c_j a constant independent of n, $j \in [\![p]\!]$, L is of order $(\lfloor \alpha_1 \rfloor,\dots,\lfloor \alpha_d \rfloor)$ and $g^{(p)} \in \Sigma(\boldsymbol{\alpha}, R)$. Then, with $\boldsymbol{h}_{\text{opt}}$ defined via $h_{k,\text{opt}} = n^{-\frac{\overline{\alpha}p}{\alpha_k(2\overline{\alpha}+d)}}$ for $k \in [\![d]\!]$ and some $C \in \mathbb{R}_+$ the resulting rate is

$$\mathbb{E}\left[(g_{\boldsymbol{h}_{\text{opt}}}^{(p)}(\boldsymbol{x})-g^{(p)}(\boldsymbol{x}))^2\right] \le C \begin{cases} n^{-1}, & \text{if } \overline{\alpha} \ge \frac{d}{2(p-1)}, \\ n^{-\frac{2\overline{\alpha}p}{2\overline{\alpha}+d}}, & \text{otherwise.} \end{cases}$$

Here, $\overline{\alpha}$ denotes the harmonic mean as given in Equation (2.22).

Finally, we deduce the regularity of $g^{(p)}$ from the regularities of the densities f_j for Hölder spaces.

Lemma 2.1.29:
If $f_j \in \Sigma((\alpha_{j,1},\dots,\alpha_{j,d}),R_j)$ for $j \in [\![p]\!]$, that is, each component density belongs to a Hölder space, and $\frac{\partial^{\lfloor \alpha_{j,k} \rfloor} f_j}{(\partial x_k)^{\lfloor \alpha_{j,k} \rfloor}}$ is integrable for all $k \in [\![d]\!]$ and $j \in [\![p]\!]$, then the convoluted density $g^{(p)}$ also belongs to a Hölder space. More precisely, there exists $R \in \mathbb{R}_+$ such that $g^{(p)} \in \Sigma(\boldsymbol{\delta}, R)$ with $\delta_k = \sum_{j\in[\![p]\!]} \lfloor \alpha_{j,k} \rfloor + \max_{j\in[\![p]\!]}(\alpha_{j,k} - \lfloor \alpha_{j,k} \rfloor)$ for $k \in [\![d]\!]$.

Proof of Lemma 2.1.29. Fix $k \in [\![d]\!]$. First, note that for all $j \in [\![p]\!]$ the function f_j admits derivatives with respect to x_k up to order $\lfloor \alpha_{j,k} \rfloor$. It follows

$$\frac{\partial^{\lfloor \alpha_{1,k} \rfloor} f_1}{(\partial x_k)^{\lfloor \alpha_{1,k} \rfloor}} * \dots * \frac{\partial^{\lfloor \alpha_{p,k} \rfloor} f_p}{(\partial x_k)^{\lfloor \alpha_{p,k} \rfloor}} = \frac{\partial^{\sum_{j\in[\![p]\!]} \lfloor \alpha_{j,k} \rfloor} g^{(p)}}{(\partial x_k)^{\sum_{j\in[\![p]\!]} \lfloor \alpha_{j,k} \rfloor}}.$$

Thus, $g^{(p)}$ admits derivatives with respect to x_k up to order $\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor = \lfloor\delta_k\rfloor$. Assume without loss of generality that

$$\max_{j\in[\![p]\!]}(\alpha_{j,k} - \lfloor\alpha_{j,k}\rfloor) = \alpha_{1,k} - \lfloor\alpha_{1,k}\rfloor.$$

Then, it holds that $\delta_k - \lfloor\delta_k\rfloor = \alpha_{1,k} - \lfloor\alpha_{1,k}\rfloor$ and it follows for $\boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ and $\boldsymbol{x}' = (x_1, \dots, x_{k-1}, x'_k, x_{k+1}, \dots, x_d) \in \mathbb{R}^d$ that

$$\begin{aligned}
&\left|\frac{\partial^{\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor} g^{(p)}}{(\partial x_k)^{\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor}}(\boldsymbol{x}) - \frac{\partial^{\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor} g^{(p)}}{(\partial x_k)^{\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor}}(\boldsymbol{x}')\right| \\
&= \left|\int_{\mathbb{R}^d}\left(\frac{\partial^{\lfloor\alpha_{1,k}\rfloor} f_1}{(\partial x_k)^{\lfloor\alpha_{1,k}\rfloor}}(\boldsymbol{x}-\boldsymbol{u}) - \frac{\partial^{\lfloor\alpha_{1,k}\rfloor} f_1}{(\partial x_k)^{\lfloor\alpha_{1,k}\rfloor}}(\boldsymbol{x}'-\boldsymbol{u})\right)\frac{\partial^{\sum_{j=2}^{p}\lfloor\alpha_{j,k}\rfloor}(f_2 * \cdots * f_p)}{(\partial x_k)^{\sum_{j=2}^{p}\lfloor\alpha_{j,k}\rfloor}}(\boldsymbol{u})d\lambda^d(\boldsymbol{u})\right| \\
&\leq R_1|x_k - x'_k|^{\alpha_{1,k}-\lfloor\alpha_{1,k}\rfloor}\int_{\mathbb{R}^d}\left|\frac{\partial^{\sum_{j=2}^{p}\lfloor\alpha_{j,k}\rfloor}(f_2 * \cdots * f_p)}{(\partial x_k)^{\sum_{j=2}^{p}\lfloor\alpha_{j,k}\rfloor}}(\boldsymbol{u})\right| d\lambda^d(\boldsymbol{u}) \\
&\leq R_1|x_k - x'_k|^{\delta_k-\lfloor\delta_k\rfloor}\prod_{j=2}^{p}\left\|\frac{\partial^{\lfloor\alpha_{j,k}\rfloor} f_j}{(\partial x_k)^{\lfloor\alpha_{j,k}\rfloor}}\right\|_{\mathbb{L}^1_d}
\end{aligned}$$

To get a constant not depending on dimension k, we upper bound by the corresponding terms by the maximum over all $k \in [\![d]\!]$, i.e.,

$$\begin{aligned}
&\left|\frac{\partial^{\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor} g^{(p)}}{(\partial x_k)^{\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor}}(\boldsymbol{x}) - \frac{\partial^{\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor} g^{(p)}}{(\partial x_k)^{\sum_{j\in[\![p]\!]}\lfloor\alpha_{j,k}\rfloor}}(\boldsymbol{x}')\right| \\
&\leq (\max_{j\in[\![p]\!]} R_j)|x_k - x'_k|^{\delta_k-\lfloor\delta_k\rfloor}\max_{k\in[\![d]\!]}\prod_{j\in[\![p]\!]}\left(\left\|\frac{\partial^{\lfloor\alpha_{j,k}\rfloor} f_j}{(\partial x_k)^{\lfloor\alpha_{j,k}\rfloor}}\right\|_{\mathbb{L}^1_d} \vee 1\right) \\
&= R|x_k - x'_k|^{\delta_k-\lfloor\delta_k\rfloor},
\end{aligned}$$

where $R := (\max_{j\in[\![p]\!]} R_j)\max_{k\in[\![d]\!]}\prod_{j\in[\![p]\!]}\left(\left\|\frac{\partial^{\lfloor\alpha_{j,k}\rfloor} f_j}{(\partial x_k)^{\lfloor\alpha_{j,k}\rfloor}}\right\|_{\mathbb{L}^1_d} \vee 1\right)$. Consequently, $g^{(p)} \in \Sigma(\boldsymbol{\delta}, R)$. This concludes the proof. □

2.1.5 Bandwidth selection with the PCO-method

The optimal choice of the bandwidth parameter $\boldsymbol{h}_{\text{opt}}$ for the Kernel estimator (2.11) depends for both the integrated and point-wise quadratic risk, see Proposition 2.1.28 and Proposition 2.1.19, on the regularity parameter of the density $g^{(p)}$ which is unknown in practice. A natural next question is to find a data-driven approach to select these parameters. From now on in this chapter, we focus on the integrated quadratic risk.

We propose to use the approach presented in Lacour et al. (2017), called Penalized Comparison to Overfitting (PCO) method. However, theoretical results are lengthy and since our small simulation study (see Section 2.3 for more details) suggests that the performance of the model selection method for projection estimation proposed in Section 2.2.4 is slightly better, we show upper bounds for the integrated squared risk only for the model selection. We outline the PCO-method adapted to the setting of this chapter for $d = 1$. Recall that we have the standard bias-variance decomposition

$$\mathbb{E}\left[\|\widehat{g_h^{(p)}} - g^{(p)}\|_{\mathbb{L}^2}^2\right] = \mathbb{E}\left[\|\widehat{g_h^{(p)}} - g_h^{(p)}\|_{\mathbb{L}^2}^2\right] + \|g_h^{(p)} - g^{(p)}\|_{\mathbb{L}^2}^2 =: v_h + b_h.$$

It is natural to propose a criterion of the form $\mathrm{Crit}(h) = \hat{b}_h + \hat{v}_h$. We have that

$$v_h \leq \frac{2\|L\|_{\mathbb{L}^2}^2}{hn_1 \cdots n_p} + C3^{2(p-2)}p \sum_{j\in[\![p]\!]} \frac{1}{n_j}$$

where $C := 4\pi\|\mathcal{F}[L]\|_{\mathbb{L}^\infty}^2 \max_{j\in[\![p]\!]} \|f_j\|_{\mathbb{L}^2}^2$. Consequently, the second summand does not depend on the bandwidth parameter h and we also use that

$$\hat{v}_h = \frac{\kappa\|L\|_{\mathbb{L}^2}^2}{hn_1 \cdots n_p} = \frac{\kappa\|L_h\|_{\mathbb{L}^2}^2}{n_1 \cdots n_p}$$

with tuning parameter κ. We choose $\kappa = 1$ due to the oracle inequality and the remark following in Lacour et al. (2017). Let $\mathcal{H}$ denote a suitable grid of bandwidths. For the estimation of the bias, assume that $h_{\min}$, the minimum of the bandwidths in the grid $\mathcal{H}$, is very small. Plugging $\widehat{g_{h_{\min}}^{(p)}}$ and $\widehat{g_h^{(p)}}$ in the bias term b_h for $g_h^{(p)}$ and $g^{(p)}$, respectively, introduces an additional bias. Consequently, a corrected term is proposed, i.e.,

$$b_h \approx \|\widehat{g_{h_{\min}}^{(p)}} - \widehat{g_h^{(p)}}\|_{\mathbb{L}^2}^2 - \frac{\|L_{h_{\min}} - L_h\|_{\mathbb{L}^2}^2}{n_1 \cdots n_p}.$$

This finally yields the criterion

$$\mathrm{Crit}(h) = \|\widehat{g_{h_{\min}}^{(p)}} - \widehat{g_h^{(p)}}\|_{\mathbb{L}^2}^2 + \frac{2\langle L_h, L_{h_{\min}}\rangle_{\mathbb{L}^2}}{n_1 \cdots n_p},$$

and the estimator $\widehat{h} = \arg\min_{h\in\mathcal{H}} \mathrm{Crit}(h)$ for the bandwidth. A data-driven estimator of $g^{(p)}$ is given by $\widehat{g_{\widehat{h}}^{(p)}}$.

This concludes the theoretical discussion of the kernel estimator. In the next section, we consider an alternative approach for the estimation of the convolution $g^{(p)}$, namely, a projection method.

2.2 Projection estimator

We propose now a projection strategy for the estimation of the convolution, first of two one-dimensional densities in Section 2.2.1. For the generalization, we introduce the Hermite basis in Section 2.2.2 and give important properties. Subsequently, we extend the method to the case of several multivariate densities in Section 2.2.3. We show upper bounds on the integrated quadratic risks. Finally, we study a model selection method in Section 2.2.4.

2.2.1 Convolution of the densities of two independent univariate variables

Let us explain the estimation method in the case of two independent one-dimensional random variables, given i.i.d. observations $(X_k^{(1)})_{k\in[\![n]\!]}$ of $X^{(1)} \sim f_1$ and $(X_k^{(2)})_{k\in[\![n]\!]}$ of $X^{(2)} \sim f_2$. The goal of this section, thus, is the estimation of $g := g^{(2)} = f_1 * f_2$.

Assume that $g \in \mathbb{L}^2$. Given an orthonormal basis $(\varphi_\ell)_{\ell \geq 0}$ of $\mathbb{L}^2$ we write

$$g = \sum_{\ell \geq 0} a_\ell \varphi_\ell, \quad a_\ell := \langle g, \varphi_\ell \rangle_{\mathbb{L}^2}.$$

Further, note that since the basis functions are assumed to be orthonormal it holds

$$\|g\|_{\mathbb{L}^2}^2 = \sum_{\ell_1 \geq 0, \ell_2 \geq 0} a_{\ell_1} \overline{a_{\ell_2}} \int_{\mathbb{R}} \varphi_{\ell_1}(t) \overline{\varphi_{\ell_2}}(t) d\lambda(t) = \sum_{\ell \geq 0} |a_\ell|^2. \tag{2.25}$$

That is, the $\mathbb{L}^2$ norm of g is equal to the sum of the squared coefficients of its development in the given basis. This property will be used repeatedly throughout this chapter. We define

$$B(m) := \sup_{x \in \mathbb{R}} \sum_{\ell=0}^{m-1} |\varphi_\ell|^2(x).$$

This term depends on the specific choice of basis and is discussed below in Remark 2.2.3 in more detail. In addition, let g_m denote the orthogonal projection of g on $S_m := \text{vect}\{\varphi_0, \dots, \varphi_{m-1}\}$ the linear space spanned by the first $m \in \mathbb{N}$ basis functions, i.e.,

$$g_m = \sum_{\ell=0}^{m-1} a_\ell \varphi_\ell.$$

To build an estimator of g, we estimate the first coefficients of its development in the basis and get, for $m \in \mathbb{N}$ the proposal

$$\widehat{g_m} := \sum_{\ell=0}^{m-1} \widehat{a_\ell} \varphi_\ell, \quad \widehat{a_\ell} := \frac{1}{n^2} \sum_{k_1 \in [\![n]\!]} \sum_{k_2 \in [\![n]\!]} \overline{\varphi_\ell}(X_{k_1}^{(1)} + X_{k_2}^{(2)}). \tag{2.26}$$

As $\mathbb{E}[\widehat{a_\ell}] = \langle g, \varphi_\ell \rangle_{\mathbb{L}^2} = a_\ell$, this provides an unbiased estimator of g_m. Note that this estimator is different from the convolution of two projection density estimators. However, analogously to the corresponding projection estimator for density estimation, the estimator given in Equation (2.26) is equal to the corresponding minimum contrast estimator which is motivated by minimizing the term $\|g_m - h\|_{\mathbb{L}^2}^2$ over S_m. More precisely, for the contrast defined for $h \in S_m$ by

$$\gamma_n(h) := \|h\|_{\mathbb{L}^2}^2 - \frac{2}{n^2} \sum_{k_1 \in [\![n]\!]} \sum_{k_2 \in [\![n]\!]} \overline{h}(X_{k_1}^{(1)} + X_{k_2}^{(2)}),$$

the minimum contrast estimator is defined as

$$\widetilde{g_m} = \arg\min_{h \in S_m} \gamma_n(h). \tag{2.27}$$

Lemma 2.2.1:
The projection estimator $\widehat{g_m}$ given in Equation (2.26) and the minimum contrast estimator $\widetilde{g_m}$ given in Equation (2.27) are equivalent.

Proof of Lemma 2.2.1. The calculations are standard, see for example Comte (2017), and we give them for the sake of completeness. Indeed, since $h \in S_m$ we write $h = \sum_{\ell=0}^{m-1} a_\ell \varphi_\ell$ and identify the function with its coefficients $(a_\ell)_{0\leq \ell \leq m-1}$. We rewrite the contrast using Equation (2.25) as

$$\gamma_n(a_0, \dots, a_{m-1}) = \sum_{\ell=0}^{m-1} |a_\ell|^2 - \frac{2}{n^2} \sum_{\ell=0}^{m-1} a_\ell \sum_{k_1 \in [\![n]\!]} \sum_{k_2 \in [\![n]\!]} \overline{\varphi_\ell}(X_{k_1}^{(1)} + X_{k_2}^{(2)}).$$

Taking the derivative with respect to one coefficient a_{ℓ_0}, $\ell_0 \in \{0, \dots, m-1\}$ we get

$$\frac{\partial \gamma_n}{\partial a_{\ell_0}}(a_0, \dots, a_{m-1}) = 2a_{\ell_0} - \frac{2}{n^2} \sum_{k_1 \in [\![n]\!]} \sum_{k_2 \in [\![n]\!]} \overline{\varphi_{\ell_0}}(X_{k_1}^{(1)} + X_{k_2}^{(2)}).$$

Further, the function $(a_0, \dots, a_{m-1}) \mapsto \gamma_n(a_0, \dots, a_{m-1})$ has Hessian matrix $2\,\mathrm{Id}_m$, where Id_m denotes the $(m \times m)$-identity matrix. $2\,\mathrm{Id}_m$ is clearly positive definite. Consequently, the function admits a unique minimum at

$$\tilde{a}_\ell = \frac{1}{n^2} \sum_{k_1 \in [\![n]\!]} \sum_{k_2 \in [\![n]\!]} \overline{\varphi_\ell}(X_{k_1}^{(1)} + X_{k_2}^{(2)}) = \widehat{a}_\ell.$$

It follows that $\widehat{g_m} = \widetilde{g_m}$, which concludes the proof. □

Recall that from $f_1, f_2 \in \mathbb{L}^2$ it follows that $g \in \mathbb{L}^\infty$, see Section 2.1.1. We obtain the following bound for the quadratic integrated risk of the projection estimator $\widehat{g_m}$.

Proposition 2.2.2:
If $f_1, f_2 \in \mathbb{L}^2$, it holds for the estimator $\widehat{g_m}$ defined in Equation (2.26) that

$$\mathbb{E}\left[\|\widehat{g_m} - g\|_{\mathbb{L}^2}^2\right] \leq \|g_m - g\|_{\mathbb{L}^2}^2 + \frac{B(m) \wedge m\|g\|_{\mathbb{L}^\infty}}{n^2} + \frac{\|f_1\|_{\mathbb{L}^2}^2 + \|f_2\|_{\mathbb{L}^2}^2}{n}.$$

Remark 2.2.3:
Proposition 2.2.2 holds for any orthonormal basis of $\mathbb{L}^2$. However, if $g \in \mathbb{L}^2(\mathbb{1}_A)$ for $A \subseteq \mathbb{R}$, results can be adapted for corresponding orthonormal bases. The impact of the choice is reflected in the term $B(m)$. For example, the trigonometric basis satisfies $B(m) = m$ (Comte (2017), Remark 1.4, Chap. 2), for the Legendre polynomial basis $B(m) \sim m^2$ (Cohen et al. (2013), p. 831) and for the Hermite basis, it holds $B(m) \sim \sqrt{m}$ (Comte and Lacour (2023), Lemma 1 and Section 2.2.2). Consequently, for the Hermite the order of the variance term improves to $\sqrt{m}/n^2$. This is coherent with the influence of the dimension of the Hermite space on the order of the bias term (see Belomestny et al. (2019)).

Similarly to the kernel estimator, we cannot apply standard methods for the variance since the n^2 variables involved in the definition of $\widehat{a}_\ell$ are not independent but direct computation is still possible, using again Lemma 2.1.5.

Proof of Proposition 2.2.2. Using the squared bias variance decomposition it remains to show an upper bound for the variance. First, we have using Equation (2.25) that

$$\mathbb{E}\left[\|\widehat{g_m} - g_m\|_{\mathbb{L}^2}^2\right] = \sum_{\ell=0}^{m-1} \mathbb{E}[|\widehat{a}_\ell - a_\ell|^2] = \sum_{\ell=0}^{m-1} \mathrm{var}\left(\frac{1}{n^2} \sum_{j \in [\![n]\!]} \sum_{k \in [\![n]\!]} \overline{\varphi_\ell}(X_j^{(1)} + X_k^{(2)})\right)$$

Consequently, using Lemma 2.1.5 with $F = \overline{\varphi_\ell} \in \mathbb{L}^2$, we have that

$$\begin{aligned}
&\mathbb{E}\left[\|\widehat{g_m} - g_m\|_{\mathbb{L}^2}^2\right] \\
&= \frac{1}{n}\sum_{\ell=0}^{m-1} \frac{n-1}{n} \Big(\operatorname{cov}\left(\overline{\varphi_\ell}(X_1^{(1)} + X_1^{(2)}), \overline{\varphi_\ell}(X_1^{(1)} + X_2^{(2)})\right) \\
&\qquad\qquad + \operatorname{cov}\left(\overline{\varphi_\ell}(X_1^{(1)} + X_1^{(2)}), \overline{\varphi_\ell}(X_2^{(1)} + X_1^{(2)})\right)\Big) \\
&+ \frac{1}{n^2}\sum_{\ell=0}^{m-1} \operatorname{var}(\overline{\varphi_\ell}(X_1^{(1)} + X_1^{(2)})) = T_1 + T_2 + T_3.
\end{aligned}$$

For the third summand, using that the basis is orthonormal and that g is a density, it holds that

$$T_3 \leq \frac{1}{n^2}\sum_{\ell=0}^{m-1} \mathbb{E}[|\varphi_\ell(X_1^{(1)} + X_1^{(2)})|^2] = \frac{1}{n^2}\sum_{\ell=0}^{m-1} \int_{\mathbb{R}} |\varphi_\ell(x)|^2 g(x) d\lambda(x) \leq \frac{1}{n^2}\left(B(m) \wedge \|g\|_\infty m\right).$$

For the first summand, we see that

$$\begin{aligned}
&\operatorname{cov}\left(\overline{\varphi_\ell}(X_1^{(1)} + X_1^{(2)}), \overline{\varphi_\ell}(X_1^{(1)} + X_2^{(2)})\right) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \overline{\varphi_\ell}(x+y) f_2(y) d\lambda(y)\right)\left(\int_{\mathbb{R}} \varphi_\ell(x+z) f_2(z) d\lambda(z)\right) f_1(x) d\lambda(x) \\
&\quad - \left|\int_{\mathbb{R}} \overline{\varphi_\ell}(x) g(x) d\lambda(x)\right|^2 \\
&= \int_{\mathbb{R}} \left|\int_{\mathbb{R}} \overline{\varphi_\ell}(u) f_2(u-x) d\lambda(u)\right|^2 f_1(x) d\lambda(x) - \left|\int_{\mathbb{R}} \overline{\varphi_\ell}(x) g(x) d\lambda(x)\right|^2.
\end{aligned}$$

Denoting by Π_{S_m} the orthogonal projection on S_m we get that

$$\begin{aligned}
nT_1 &\leq \sum_{\ell=0}^{m-1} \operatorname{cov}\left(\overline{\varphi_\ell}(X_1^{(1)} + X_1^{(2)}), \overline{\varphi_\ell}(X_1^{(1)} + X_2^{(2)})\right) \\
&\leq \sum_{\ell=0}^{m-1} \int_{\mathbb{R}} \left|\int_{\mathbb{R}} \overline{\varphi_\ell}(x) f_2(u-x) d\lambda(u)\right|^2 f_1(x) d\lambda(x) \\
&= \int_{\mathbb{R}} \sum_{\ell=0}^{m-1} |\langle f_2(\cdot - x), \varphi_\ell\rangle|^2 f_1(x) d\lambda(x) = \int_{\mathbb{R}} \|\Pi_{S_m} f_2(\cdot - x)\|_{\mathbb{L}^2}^2 f_1(x) d\lambda(x) \\
&\leq \int_{\mathbb{R}} \|f_2(\cdot - x)\|_{\mathbb{L}^2}^2 f_1(x) d\lambda(x) = \int_{\mathbb{R}} f_2^2(v) d\lambda(v) \int_{\mathbb{R}} f_1(x) d\lambda(x) = \int_{\mathbb{R}} f_2^2(v) d\lambda(v).
\end{aligned}$$

Analogously, one can bound

$$nT^2 \leq \sum_{\ell=0}^{m-1} \operatorname{cov}\left(\overline{\varphi_\ell}(X_1^{(1)} + X_1^{(2)}), \overline{\varphi_\ell}(X_2^{(1)} + X_1^{(2)})\right) \leq \int_{\mathbb{R}} f_1^2(v) d\lambda(v).$$

With this, it follows that

$$T_1 + T_2 \leq \frac{1}{n}\left(\int_{\mathbb{R}} f_2^2(v) d\lambda(v) + \int_{\mathbb{R}} f_1^2(v) d\lambda(v)\right),$$

which concludes the proof. □

2.2.2 Hermite basis

To extend the definition of the projection estimator to the convolution of more than two variables, it is convenient to rely on a basis admitting a regular Fourier transform. This is why we develop extensions relying on the Hermite basis, and, thus, start by giving its definition and useful properties.

We start with the orthonormal Hermite basis for $A = \mathbb{R}$, and then build a basis for $\mathbb{R}^d$ by tensorization. The Hermite basis $(\varphi_\ell)_{\ell\geq 0}$ is a basis on $\mathbb{L}^2$ defined from the Hermite polynomials $(H_\ell)_{\ell\geq 0}$, given by

$$H_\ell(x) := (-1)^\ell \exp(x^2)\frac{\partial^\ell}{(\partial x)^\ell}\exp(-x^2).$$

Then, for $\ell \geq 0$, the elements of the Hermite basis are defined by

$$\varphi_\ell(x) := c_\ell H_\ell(x)\exp(-x^2/2), \quad c_\ell := (2^\ell \ell! \sqrt{\pi})^{-1/2}. \tag{2.28}$$

The Hermite basis has the following two properties which will be used in the proofs. Firstly, the Fourier transform of each basis function verifies for $\ell \geq 0$

$$\mathcal{F}[\varphi_\ell] = \sqrt{2\pi} i^\ell \varphi_\ell. \tag{2.29}$$

Consequently, the Fourier transform is equal to the basis function itself up to a constant. Secondly, we have according to Askey and Wainger (1965) that there exist numerical constants $\xi, \kappa_\varphi \in \mathbb{R}_+$ such that for all $\ell \in \mathbb{N}_0$ it holds

$$|\varphi_\ell(x)| < \kappa_\varphi \exp(-\xi x^2), \quad |x| \geq \sqrt{2\ell+1}. \tag{2.30}$$

The last property we use to control the $\mathbb{L}^1_d$–norm of the basis functions. More precisely, we use the following lemma.

Lemma 2.2.4:
For $m \in \mathbb{N}$ and $\ell \in \{0, \ldots, m_k\}$ it holds that

$$\int_{|t|>\sqrt{2m+1}} |\varphi_\ell(t)| d\lambda(t) \leq \frac{\kappa_\varphi \exp(-\xi(2m+1))}{\xi\sqrt{2m+1}} \tag{2.31}$$

with ξ and κ_φ from Equation (2.30) and

$$\left(\int_{\mathbb{R}} |\varphi_\ell(t)|\, d\lambda(t)\right)^2 \leq 6\sqrt{2m+1}. \tag{2.32}$$

Proof of Lemma 2.2.4. First note that for $A \geq 1$ it holds that

$$\int_A^\infty \exp(-cx^2) d\lambda(x) = \int_A^\infty x\exp(-cx^2)\frac{1}{x} d\lambda(x) \leq \frac{1}{A}\int_A^\infty x\exp(-cx^2)d\lambda(x) = \frac{\exp(-cA^2)}{2cA}.$$

Together with Equation (2.30), it follows for $\ell \in \{0, \ldots, m-1\}$ that

$$\int_{|t|>\sqrt{2m+1}} |\varphi_\ell(t)| d\lambda(t) \leq \int_{|t|>\sqrt{2m+1}} \kappa_\varphi \exp(-\xi t^2) d\lambda(t) \leq \frac{\kappa_\varphi \exp(-\xi(2m+1))}{\xi\sqrt{2m+1}}.$$

In addition, we have that

$$\begin{aligned}\left(\int_{\mathbb{R}} |\varphi_\ell(t)|\, d\lambda(t)\right)^2 &= \left(\int_{|t|<\sqrt{2m+1}} |\varphi_\ell(t)|\, d\lambda(t) + \int_{|t|>\sqrt{2m+1}} |\varphi_\ell(t)|\, d\lambda(t)\right)^2 \\ &\le 2\left(\int_{\mathbb{R}} \mathbb{1}_{\{|t|<\sqrt{2m+1}\}} d\lambda(t) \int_{\mathbb{R}} |\varphi_\ell(t)|^2\, d\lambda(t)\right) + \frac{2\exp(-2\xi(2m+1))}{2m+1} \\ &= 2\left(2\sqrt{2m+1} + \frac{\exp(-2\xi(2m+1))}{2m+1}\right) \le 6\sqrt{2m+1}.\end{aligned}$$

This concludes the proof. □

We refer to Sacko (2020) for more details on density estimation with the Hermite basis. Using the Hermite basis on $\mathbb{L}^2$, we build a basis on $\mathbb{L}^2_d$. That is, for $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_d) \in \mathbb{N}_0^d$, define

$$\phi_{\boldsymbol{\ell}}(\boldsymbol{x}) := \phi_{\ell_1,\ldots,\ell_d}(\boldsymbol{x}) := (\varphi_{\ell_1} \otimes \cdots \otimes \varphi_{\ell_d})(\boldsymbol{x}) = \varphi_{\ell_1}(x_1)\cdots\varphi_{\ell_d}(x_d). \tag{2.33}$$

Note that due to Equation (2.29) we have for the Hermite basis on $\mathbb{L}^2_d$ for the Fourier transform for all $\ell \in \mathbb{N}_0^d$ it holds

$$\begin{aligned}\mathcal{F}[\phi_{\boldsymbol{\ell}}](\boldsymbol{x}) = \mathcal{F}[\phi_{\ell_1,\ldots,\ell_d}](\boldsymbol{x}) &= \mathcal{F}[\varphi_{\ell_1}](x_1)\cdots\mathcal{F}[\varphi_{\ell_d}](x_d) = (\sqrt{2\pi})^d i^{\ell_1+\ldots+\ell_d}\varphi_{\ell_1}(x_1)\cdots\varphi_{\ell_d}(x_d) \\ &= (\sqrt{2\pi})^d i^{\ell_1+\ldots+\ell_d}\phi_{\ell_1,\ldots,\ell_d}(\boldsymbol{x}) = (\sqrt{2\pi})^d i^{\ell_1+\ldots+\ell_d}\phi_{\boldsymbol{\ell}}(\boldsymbol{x}).\end{aligned} \tag{2.34}$$

Further, for the Hermite basis, it holds for all $\boldsymbol{t} \in \mathbb{R}^d$ and $\boldsymbol{\ell} \in \mathbb{N}_0^d$ that $\phi_{\boldsymbol{\ell}}(\boldsymbol{t}) \in \mathbb{R}$, and thus, complex conjugation is omitted when applied to these functions.

2.2.3 Convolution of the densities of p independent multivariate variables

We now extend the projection estimator proposed in Section 2.2.1 to the multivariate multiple density case. Recall, the goal is to estimate $g^{(p)} : \mathbb{R}^d \to \mathbb{R}$ defined by Equation (2.1) from independent vector observations which are all d-dimensional but with possibly different sample sizes, i.e., for $j \in [\![p]\!]$, we observe $\boldsymbol{X}_k^{(j)} = (X_{k,1}^{(j)}, \ldots, X_{k,d}^{(j)}) \in \mathbb{R}^d$, with $\boldsymbol{X}_k^{(j)} \sim f_j$ i.i.d., for $k \in [\![n_j]\!]$ and independent for $j \in [\![p]\!]$.

Assume that $g^{(p)} \in \mathbb{L}^2_d$. Given the Hermite orthonormal basis $(\phi_{\boldsymbol{\ell}})_{\boldsymbol{\ell} \ge \mathbb{N}_0^d}$ of $\mathbb{L}^2_d$ introduced in the previous section, we write

$$g^{(p)} = \sum_{\boldsymbol{\ell}\ge 0} a_{\boldsymbol{\ell}}\phi_{\boldsymbol{\ell}} = \sum_{\ell_1\ge 0}\cdots\sum_{\ell_d\ge 0} a_{\ell_1,\ldots,\ell_d}\phi_{\ell_1,\ldots,\ell_d}$$

with $a_{\boldsymbol{\ell}} = a_{\ell_1,\ldots,\ell_d} := \langle g^{(p)}, \phi_{\ell_1,\ldots,\ell_d}\rangle_{\mathbb{L}^2_d}$. Moreover, for $\boldsymbol{m} = (m_1, \ldots, m_d) \in \mathbb{N}_0^d$ let

$$g^{(p)}_{\boldsymbol{m}} = \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-\boldsymbol{1}} a_{\boldsymbol{\ell}}\phi_{\boldsymbol{\ell}} = \sum_{\ell_1=0}^{m_1-1}\cdots\sum_{\ell_d=0}^{m_d} a_{\ell_1,\ldots,\ell_d}\phi_{\ell_1,\ldots,\ell_d}$$

denote the orthogonal projection of $g^{(p)}$ on

$$S_{\boldsymbol{m}} := \mathrm{vect}\{\phi_{\boldsymbol{\ell}} : \forall k \in [\![d]\!] : \ell_k \le m_k - 1\}.$$

For $\boldsymbol{m} = (m_1, \ldots, m_d) \in \mathbb{N}_0^d$ and $\boldsymbol{n} = (n_1, \ldots, n_p) \in \mathbb{N}^d$, this motivates analogously to Equation (2.26) the estimator

$$\widehat{g_{\boldsymbol{m}}^{(p)}} := \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} \widehat{a_{\boldsymbol{\ell}}} \phi_{\boldsymbol{\ell}} \tag{2.35}$$

$$\widehat{a_{\boldsymbol{\ell}}} := \frac{1}{n_1 \cdots n_p} \sum_{k_1 \in [\![n_1]\!]} \cdots \sum_{k_p \in [\![n_p]\!]} \phi_{\boldsymbol{\ell}}(\boldsymbol{X}_{k_1}^{(1)} + \ldots + \boldsymbol{X}_{k_p}^{(p)}). \tag{2.36}$$

Note that we rewrite the coefficients using that $\phi_{\boldsymbol{\ell}} = \mathcal{F}^\dagger[\mathcal{F}[\phi_{\boldsymbol{\ell}}]]$ for all $\boldsymbol{\ell} \in \mathbb{N}_0^d$. More precisely, we have that

$$\begin{aligned}
\widehat{a_{\boldsymbol{\ell}}} &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \frac{1}{n_1 \cdots n_p} \sum_{k_1 \in [\![n_1]\!]} \cdots \sum_{k_p \in [\![n_p]\!]} e^{-i\langle \boldsymbol{u}, \boldsymbol{X}_{k_1}^{(1)} + \ldots + \boldsymbol{X}_{k_p}^{(p)} \rangle} \mathcal{F}[\phi_{\boldsymbol{\ell}}](\boldsymbol{u}) d\lambda^d(\boldsymbol{u}) \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \frac{1}{n_1 \cdots n_p} \sum_{k_1 \in [\![n_1]\!]} \cdots \sum_{k_p \in [\![n_p]\!]} e^{i\langle \boldsymbol{u}, \boldsymbol{X}_{k_1}^{(1)} + \ldots + \boldsymbol{X}_{k_p}^{(p)} \rangle} \mathcal{F}[\phi_{\boldsymbol{\ell}}](-\boldsymbol{u}) d\lambda^d(\boldsymbol{u}) \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \prod_{j \in [\![p]\!]} \widehat{\mathcal{F}[f_p]}(\boldsymbol{u}) \overline{\mathcal{F}[\phi_{\boldsymbol{\ell}}]}(\boldsymbol{u}) d\lambda^d(\boldsymbol{u}) \\
&= \frac{1}{(2\pi)^d} \langle \widehat{\mathcal{F}[f_1]} \cdots \widehat{\mathcal{F}[f_p]}, \mathcal{F}[\phi_{\boldsymbol{\ell}}] \rangle_{\mathbb{L}_d^2},
\end{aligned}$$

where $\widehat{\mathcal{F}[f_p]}$, $j \in [\![p]\!]$, denotes the empirical transform, as introduced in Equation (2.12). Further, $\widehat{a_{\boldsymbol{\ell}}}$ are unbiased estimators of $a_{\boldsymbol{\ell}}$ for all $\boldsymbol{\ell} \in \mathbb{N}_0^d$. That is, using that the Hermite basis elements are real-valued it holds

$$\begin{aligned}
\mathbb{E}\left[\widehat{a_{\boldsymbol{\ell}}}\right] = \mathbb{E}[\phi_{\boldsymbol{\ell}}(\boldsymbol{X}_{k_1}^{(1)} + \ldots + \boldsymbol{X}_{k_p}^{(p)})] &= \int_{\mathbb{R}^d} g^{(p)}(\boldsymbol{u}) \phi_{\boldsymbol{\ell}}(\boldsymbol{u}) d\lambda^d(\boldsymbol{u}) \\
&= \langle g^{(p)}, \phi_{\boldsymbol{\ell}} \rangle_{\mathbb{L}^2} = a_{\boldsymbol{\ell}}.
\end{aligned}$$

From this follows that for all $\boldsymbol{x} \in \mathbb{R}^d$ the estimator $\widehat{g_{\boldsymbol{m}}^{(p)}}(\boldsymbol{x})$ defines an unbiased estimator of $g_{\boldsymbol{m}}^{(p)}(\boldsymbol{x})$.

In addition, analogously to the previous case, one can extend the contrast Equation (2.27) to the multivariate case with multiple densities: For $t \in S_{\boldsymbol{m}}$, define

$$\gamma_n(t) := \|t\|_{\mathbb{L}_d^2}^2 - \frac{2}{n_1 \cdots n_p} \sum_{\boldsymbol{k} \in [\![\boldsymbol{n}]\!]} t(\boldsymbol{X}_{k_1}^{(1)} + \ldots + \boldsymbol{X}_{k_p}^{(p)}).$$

One can show that the given projection estimator minimizes the contrast on $S_{\boldsymbol{m}}$, analogously to Lemma 2.2.1.

To be able to control the integrated quadratic risk of the proposed estimator, it is necessary that all basis elements $\phi_{\boldsymbol{\ell}}$, $\boldsymbol{\ell} \in \mathbb{N}_0^d$ satisfy $\mathcal{F}[\phi_{\boldsymbol{\ell}}] \in \mathbb{L}_d^1 \cap \mathbb{L}_d^2$. This is indeed that case since $\phi_{\boldsymbol{\ell}} \in \mathbb{L}_d^1 \cap \mathbb{L}_d^2$ and with Equation (2.34) the Fourier transform $\mathcal{F}[\phi_{\boldsymbol{\ell}}]$ is equal to $\phi_{\boldsymbol{\ell}}$ up to constants.

Proposition 2.2.5 (Bound on the integrated quadratic risk):
For some $C_0 \in \mathbb{R}_+$ let $\boldsymbol{m} \in \mathbb{N}^d$ satisfy

$$\frac{1}{2\xi} \log\left(\prod_{j\in[\![p]\!]} n_j\right) \leq m_k,\ \forall k \in [\![d]\!] \quad \text{and} \quad \frac{\prod_{k\in[\![d]\!]} \sqrt{2m_k+1}}{\prod_{j\in[\![p]\!]} n_j} \leq C_0. \tag{2.37}$$

If $\mathcal{F}[f_j] \in \mathbb{L}^1_d \cap \mathbb{L}^2_d$ for all $j \in [\![p]\!]$, then, the estimator $\widehat{g^{(p)}_{\boldsymbol{m}}}$ of $g^{(p)}$ defined in (2.35) with $\phi_{\boldsymbol{\ell}}$ as in (2.28)-(2.33), satisfies

$$\mathbb{E}\left[\|\widehat{g^{(p)}_{\boldsymbol{m}}} - g^{(p)}\|^2_{\mathbb{L}^2_d}\right] \leq \|g^{(p)}_{\boldsymbol{m}} - g^{(p)}\|^2_{\mathbb{L}^2_d} + C_1 \frac{\prod_{k\in[\![d]\!]} \sqrt{2m_k+1}}{\prod_{j\in[\![p]\!]} n_j} + C_2 \sum_{j\in[\![p]\!]} \frac{1}{n_j},$$

with $C_1 = \frac{8}{(2\pi)^d}$ and $C_2 = 4p3^{2(p-2)}(2\pi)^d \max_{j\in[\![p]\!]} \|f_j\|^2_{\mathbb{L}^2_d} + \frac{C_0 2d^2 6^{d-1}\kappa^2_\varphi}{\xi^2}$, with κ_φ and ξ defined in Equation (2.30).

Remark 2.2.6:
Condition (2.37) is composed of two bounds on $\boldsymbol{m}$ which are rather weak. The upper bound part ensures the boundedness of the variance and of cross terms. The lower bound part on the m_k's imposes minimal values more that log of the sample sizes and is not a strong restriction; in particular, it would allow for the optimal choice of $\boldsymbol{m}$, see Remark 2.2.7.

Proof of Proposition 2.2.5. First, since $\mathbb{E}[\widehat{g^{(p)}_{\boldsymbol{m}}}(\boldsymbol{x})] = g^{(p)}_{\boldsymbol{m}}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^d$, we see with the standard decomposition that

$$\mathbb{E}\left[\|\widehat{g^{(p)}_{\boldsymbol{m}}} - g^{(p)}\|^2_{\mathbb{L}^2_d}\right] = \|g^{(p)}_{\boldsymbol{m}} - g^{(p)}\|^2_{\mathbb{L}^2_d} + \mathbb{E}\left[\|\widehat{g^{(p)}_{\boldsymbol{m}}} - g^{(p)}_{\boldsymbol{m}}\|^2_{\mathbb{L}^2_d}\right].$$

Further, it holds that

$$\begin{aligned}\mathbb{E}\left[\|\widehat{g^{(p)}_{\boldsymbol{m}}} - g^{(p)}_{\boldsymbol{m}}\|^2\right] &= \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} \mathbb{E}\left[|\widehat{a_{\boldsymbol{\ell}}} - a_{\boldsymbol{\ell}}|^2\right] \\ &= \frac{1}{(2\pi)^{2d}} \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} \mathbb{E}\left[\left|\langle \widehat{\mathcal{F}[f_1]} \cdots \widehat{\mathcal{F}[f_p]} - \mathcal{F}[f_1]\cdots\mathcal{F}[f_p], \mathcal{F}[\phi_{\boldsymbol{\ell}}]\rangle_{\mathbb{L}^2_d}\right|^2\right].\end{aligned}$$

Recall that the Hermite basis is real-valued and consequently it holds that its complex conjugation is equal to the function itself, i.e., for all $\boldsymbol{t} \in \mathbb{R}^d$ it holds $\phi_{\boldsymbol{\ell}}(\boldsymbol{t}) = \overline{\phi_{\boldsymbol{\ell}}}(\boldsymbol{t})$. Further, from Equation (2.34) it holds that for all $\boldsymbol{t} \in \mathbb{R}^d$ it holds $\mathcal{F}[\phi_{\boldsymbol{\ell}}](\boldsymbol{t}) = (\sqrt{2\pi})^d i^{\ell_1+\ldots+\ell_d}\phi_{\boldsymbol{\ell}}(\boldsymbol{x})\phi_{\boldsymbol{\ell}}(\boldsymbol{t})$, and, thus, we obtain

$$\begin{aligned}\mathbb{E}\left[\|\widehat{g^{(p)}_{\boldsymbol{m}}} - g^{(p)}_{\boldsymbol{m}}\|^2\right] &\leq \frac{1}{(2\pi)^d} \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} \mathbb{E}\left[\left|\langle \widehat{\mathcal{F}[f_1]} \cdots \widehat{\mathcal{F}[f_p]} - \mathcal{F}[f_1]\cdots\mathcal{F}[f_p], \phi_{\boldsymbol{\ell}}\rangle_{\mathbb{L}^2_d}\right|^2\right] \\ &= \frac{1}{(2\pi)^d} \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} \mathbb{E}\left[\left|\int_{\mathbb{R}^d} \left(\prod_{j\in[\![p]\!]} \widehat{\mathcal{F}[f_j]}(\boldsymbol{t}) - \prod_{j\in[\![p]\!]} \mathcal{F}[f_j](\boldsymbol{t})\right) \phi_{\boldsymbol{\ell}}(\boldsymbol{t}) d\lambda^d(\boldsymbol{t})\right|^2\right].\end{aligned}$$

Denote

$$\widehat{\Delta\mathcal{F}[f]}(\boldsymbol{t}) := \prod_{j\in[\![p]\!]} \widehat{\mathcal{F}[f_j]}(\boldsymbol{t}) - \prod_{j\in[\![p]\!]} \mathcal{F}[f_j](\boldsymbol{t}).$$

We split the integral on the set $|\boldsymbol{t}| \leq \sqrt{2\boldsymbol{m}+1}$ (understood component wise) and its complement and bound each summand separately. More precisely, we have that

$$\mathbb{E}\left[\|\widehat{g_{\boldsymbol{m}}^{(p)}} - g_{\boldsymbol{m}}^{(p)}\|_{\mathbb{L}_d^2}^2\right] \leq \frac{2}{(2\pi)^d}(T_1 + T_2)$$

with

$$T_1 := \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} \mathbb{E}\left[\left|\int_{|\boldsymbol{t}|\leq\sqrt{2\boldsymbol{m}+1}} \left(\widehat{\Delta\mathcal{F}[f]}(\boldsymbol{t})\right) \phi_{\boldsymbol{\ell}}(t) d\lambda^d(\boldsymbol{t})\right|^2\right],$$

$$T_2 := \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} \mathbb{E}\left[\left|\int_{\mathbb{R}^d} \mathbb{1}_{\{\exists k\in[\![d]\!]:|t_k|>\sqrt{2m_k+1}\}} \left(\widehat{\Delta\mathcal{F}[f]}(\boldsymbol{t})\right) \phi_{\boldsymbol{\ell}}(\boldsymbol{t}) d\lambda^d(\boldsymbol{t})\right|^2\right].$$

For T_1 we see first see that

$$\begin{aligned} T_1 &= \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} \mathbb{E}\left[|\langle \widehat{\Delta\mathcal{F}[f]}\mathbb{1}_{\{|\cdot|\leq\sqrt{2\boldsymbol{m}+1}\}}, \phi_{\boldsymbol{\ell}}\rangle_{\mathbb{L}_d^2}|^2\right] \\ &\leq \mathbb{E}\left[\|\widehat{\Delta\mathcal{F}[f]}\mathbb{1}_{\{|\cdot|\leq\sqrt{2\boldsymbol{m}+1}\}}\|_{\mathbb{L}_d^2}^2\right] = \mathbb{E}\left[\int_{\mathbb{R}^d} \left|\widehat{\Delta\mathcal{F}[f]}(\boldsymbol{t})\right|^2 \mathbb{1}_{\{|\boldsymbol{t}|\leq\sqrt{2\boldsymbol{m}+1}\}} d\lambda^d(\boldsymbol{t})\right]. \end{aligned}$$

With Equation (2.19) it follows that

$$\begin{aligned} T_1 &\leq \mathbb{E}\left[\int_{\mathbb{R}^d} \left|\widehat{\Delta\mathcal{F}[f]}(\boldsymbol{t})\right|^2 \mathbb{1}_{\{|\boldsymbol{t}|\leq\sqrt{2\boldsymbol{m}+1}\}} d\lambda^d(\boldsymbol{t})\right] = \int_{\mathbb{R}^d} \mathbb{E}\left[\left|\widehat{\Delta\mathcal{F}[f]}(\boldsymbol{t})\right|^2\right] \mathbb{1}_{\{|\boldsymbol{t}|\leq\sqrt{2\boldsymbol{m}+1}\}} d\lambda^d(\boldsymbol{t}) \\ &= \int_{\mathbb{R}^d} \left(\frac{2}{n_1\cdots n_p} + 2p\cdot 3^{2(p-2)} \sum_{j\in[\![p]\!]} \frac{|\mathcal{F}[f_{j+1}](\boldsymbol{t})|^2}{n_j}\right) \mathbb{1}_{\{|\boldsymbol{t}|\leq\sqrt{2\boldsymbol{m}+1}\}} d\lambda^d(\boldsymbol{t}) \\ &\leq \frac{2\cdot 2^d \prod_{k\in[\![d]\!]}\sqrt{2m_k+1}}{\prod_{j\in[\![p]\!]} n_j} + 2p\cdot 3^{2(p-2)} \max_{j\in[\![p]\!]} \|\mathcal{F}[f_j]\|_{\mathbb{L}_d^2}^2 \sum_{j\in[\![p]\!]} \frac{1}{n_j}. \end{aligned} \tag{2.38}$$

For T_2, denote

$$\left(\widehat{\Delta\mathcal{F}[f]}\right)_1(\boldsymbol{t}) := \prod_{j\in[\![p]\!]} \left(\widehat{\mathcal{F}[f_j]}(\boldsymbol{t}) - \mathcal{F}[f_j](\boldsymbol{t})\right) \tag{2.39}$$

$$\left(\widehat{\Delta\mathcal{F}[f]}\right)_2(\boldsymbol{t}) := \left(\widehat{\Delta\mathcal{F}[f]}\right)(\boldsymbol{t}) - \left(\widehat{\Delta\mathcal{F}[f]}\right)_1(\boldsymbol{t}) \tag{2.40}$$

and use this to further split T_2. That is,

$$\begin{aligned} T_2 &\leq 2\sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} \mathbb{E}\left[\left|\int_{\mathbb{R}^d} \mathbb{1}_{\{\exists k\in[\![d]\!]:|t_k|>\sqrt{2m_k+1}\}} \left(\widehat{\Delta\mathcal{F}[f]}(\boldsymbol{t})\right)_1 \phi_{\boldsymbol{\ell}}(\boldsymbol{t}) d\lambda^d(\boldsymbol{t})\right|^2\right] \\ &\quad + 2\sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} \mathbb{E}\left[\left|\int_{\mathbb{R}^d} \mathbb{1}_{\{\exists k\in[\![d]\!]:|t_k|>\sqrt{2m_k+1}\}} \left(\widehat{\Delta\mathcal{F}[f]}(\boldsymbol{t})\right)_2 \phi_{\boldsymbol{\ell}}(\boldsymbol{t}) d\lambda^d(\boldsymbol{t})\right|^2\right]. \end{aligned}$$

We further upper bound

$$\begin{aligned} T_2 &\leq 2d \sum_{k\in[\![d]\!]} \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} \mathbb{E}\left[\left|\int_{\mathbb{R}^d} \mathbb{1}_{\{|t_k|>\sqrt{2m_k+1}\}} \left|\left(\widehat{\Delta\mathcal{F}[f]}(\boldsymbol{t})\right)_1\right| |\phi_{\boldsymbol{\ell}}(\boldsymbol{t})| d\lambda^d(\boldsymbol{t})\right|^2\right] \\ &\quad + 2\mathbb{E}\left[\int_{\mathbb{R}^d} \left|\left(\widehat{\Delta\mathcal{F}[f]}(\boldsymbol{t})\right)_2\right|^2 d\lambda^d(\boldsymbol{t})\right] \\ &=: T_{2,1} + T_{2,2}. \end{aligned}$$

With Remark 2.4.2 we get that

$$T_{2,2} \leq 2p3^{2(p-2)} \max_{j\in[\![p]\!]} \|\mathcal{F}[f_j]\|_{\mathbb{L}_d^2}^2 \sum_{k\in[\![p]\!]} \frac{1}{n_k}. \tag{2.41}$$

Consider $T_{2,1}$. Applying Cauchy–Schwarz inequality it holds

$$\begin{aligned} T_{2,1} &\leq 2d \sum_{k\in[\![d]\!]} \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} \mathbb{E}\Bigg[\int_{\mathbb{R}^d} \mathbb{1}_{\{|t_k|>\sqrt{2m_k+1}\}} \left|\left(\widehat{\Delta\mathcal{F}[f]}(\boldsymbol{t})\right)_1\right|^2 |\phi_{\boldsymbol{\ell}}(\boldsymbol{t})| d\lambda^d(\boldsymbol{t}) \\ &\qquad \cdot \int_{\mathbb{R}^d} \mathbb{1}_{\{|t_k|>\sqrt{2m_k+1}\}} |\phi_{\boldsymbol{\ell}}(\boldsymbol{t})| d\lambda^d(\boldsymbol{t})\Bigg] \\ &= 2d \sum_{k\in[\![d]\!]} \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} \int_{\mathbb{R}^d} \mathbb{1}_{\{|t_k|>\sqrt{2m_k+1}\}} \mathbb{E}\left[\left|\left(\widehat{\Delta\mathcal{F}[f]}(\boldsymbol{t})\right)_1\right|^2\right] |\phi_{\boldsymbol{\ell}}(\boldsymbol{t})| d\lambda^d(\boldsymbol{t}) \\ &\qquad \cdot \int_{\mathbb{R}^d} \mathbb{1}_{\{|t_k|>\sqrt{2m_k+1}\}} |\phi_{\boldsymbol{\ell}}(\boldsymbol{t})| d\lambda^d(\boldsymbol{t}). \end{aligned}$$

With Equation (2.18), we get due to the independence of samples that

$$\mathbb{E}\left[\left|\left(\widehat{\Delta\mathcal{F}[f]}(\boldsymbol{t})\right)_1\right|^2\right] = \prod_{j\in[\![p]\!]} \mathbb{E}\left[\left|\widehat{\mathcal{F}[f_j]}(\boldsymbol{t}) - \mathcal{F}[f_j](\boldsymbol{t})\right|^2\right] \leq \prod_{j\in[\![p]\!]} \frac{1}{n_j}.$$

From this follows for $T_{2,1}$ that

$$T_{2,1} \leq 2d\Big(\prod_{j\in[\![p]\!]} \frac{1}{n_j}\Big) \sum_{k\in[\![d]\!]} \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} \left(\int_{\mathbb{R}^d} \mathbb{1}_{\{|t_k|>\sqrt{2m_k+1}\}} |\phi_{\boldsymbol{\ell}}(\boldsymbol{t})| d\lambda^d(\boldsymbol{t})\right)^2. \tag{2.42}$$

Recall that we have defined for $\ell \in \mathbb{N}_0^d$ the Hermite basis element $\phi_{\boldsymbol{\ell}}$ on $\mathbb{R}^d$ as a product of basis elements of the Hermite basis on $\mathbb{R}$, i.e., $\phi_{\boldsymbol{\ell}}(\boldsymbol{x}) = \varphi_{\ell_1}(x_1)\cdots\varphi_{\ell_d}(x_d)$ for $\boldsymbol{x} \in \mathbb{R}^d$, see Equation (2.33). With this we rewrite the integral in the last inequality for $k \in [\![d]\!]$ and $\ell \in \mathbb{N}_0^d$ as

$$\int_{\mathbb{R}^d} \mathbb{1}_{\{|t_k|>\sqrt{2m_k+1}\}} |\phi_{\boldsymbol{\ell}}(\boldsymbol{t})| d\lambda^d(\boldsymbol{t}) = \int_{\mathbb{R}} \mathbb{1}_{\{|t_k|>\sqrt{2m_k+1}\}} |\varphi_{\ell_k}(t_k)| d\lambda(t_k) \prod_{\tilde{k}\in[\![d]\!],\tilde{k}\neq k} \int_{\mathbb{R}} |\varphi_{\tilde{k}}(t_{\tilde{k}})| d\lambda(t_{\tilde{k}}).$$

Using this equality and with Equation (2.31) and Equation (2.32) of Lemma 2.2.4 we obtain

$$\begin{aligned}
&\left(\int_{\mathbb{R}^d} \mathbb{1}_{\{|t_k|>\sqrt{2m_k+1}\}} |\phi_{\boldsymbol{\ell}}(\boldsymbol{t})| d\lambda^d(\boldsymbol{t})\right)^2 \\
&= \left(\int_{\mathbb{R}} \mathbb{1}_{\{|t_k|>\sqrt{2m_k+1}\}} |\varphi_{\ell_k}(t_k)| d\lambda(t_k)\right)^2 \prod_{\tilde{k}\in[\![d]\!],\tilde{k}\neq k} \left(\int_{\mathbb{R}} |\varphi_{\tilde{k}}(t_{\tilde{k}})| d\lambda(t_{\tilde{k}})\right)^2 \\
&\leq \frac{\kappa_\varphi^2 \exp(-2\xi(2m_k+1))}{\xi^2(2m_k+1)} \prod_{\tilde{k}\in[\![d]\!],\tilde{k}\neq k} 6\sqrt{2m_{\tilde{k}}+1} \\
&\leq \frac{\kappa_\varphi^2 6^{d-1}\exp(-4\xi m_k)}{\xi^2} \prod_{\tilde{k}\in[\![d]\!]} \sqrt{2m_{\tilde{k}}+1}.
\end{aligned}$$

Inserting this calculation into Equation (2.42) we get that

$$T_{2,1} \leq 2d\Big(\prod_{j\in[\![p]\!]} \frac{1}{n_j}\Big) \sum_{k\in[\![d]\!]} \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} \frac{\kappa_\varphi^2 6^{d-1}\exp(-4\xi m_k)}{\xi^2} \prod_{\tilde{k}\in[\![d]\!]} \sqrt{2m_{\tilde{k}}+1}.$$

Consequently, with $\boldsymbol{m}$ satisfying Equation (2.37) with $C_0 \in \mathbb{R}_+$ it follows

$$\begin{aligned}
T_{2,1} &\leq \frac{2d6^{d-1}\kappa_\varphi^2}{\xi^2} \frac{\prod_{\tilde{k}\in[\![d]\!]} \sqrt{2m_{\tilde{k}}+1}}{\prod_{j\in[\![p]\!]} n_j} \sum_{k\in[\![d]\!]} \exp\Big(-2\log\Big(\prod_{j\in[\![p]\!]} n_j\Big)\Big) \\
&\leq \frac{C_0 2d^2 6^{d-1}\kappa_\varphi^2}{\xi^2} \prod_{j\in[\![p]\!]} \frac{1}{n_j^2},
\end{aligned} \tag{2.43}$$

Putting the bounds Equations (2.38), (2.41) and (2.43) together, it follows that

$$\begin{aligned}
&\mathbb{E}\left[\|\widehat{g_{\boldsymbol{m}}^{(p)}} - g_{\boldsymbol{m}}^{(p)}\|_{\mathbb{L}^2_d}^2\right] \leq \frac{2}{(2\pi)^d}(T_1 + T_{2,2} + T_{2,1}) \\
&\leq \frac{2}{(2\pi)^d}\Big(\frac{4\prod_{k\in[\![d]\!]}\sqrt{2m_k+1}}{\prod_{j\in[\![p]\!]} n_j} + 2p\cdot 3^{2(p-2)} \max_{j\in[\![p]\!]} \|\mathcal{F}[f_j]\|_{\mathbb{L}^2_d}^2 \sum_{j\in[\![p]\!]} \frac{1}{n_j} \\
&\qquad + 2p3^{2(p-2)} \max_{j\in[\![p]\!]} \|\mathcal{F}[f_j]\|_{\mathbb{L}^2_d}^2 \sum_{j\in[\![p]\!]} \frac{1}{n_j} + \frac{C_0 2d^2 6^{d-1}\kappa_\varphi^2}{\xi^2} \prod_{j\in[\![p]\!]} \frac{1}{n_j^2}\Big) \\
&\leq C_1 \frac{\prod_{k\in[\![d]\!]}\sqrt{2m_k+1}}{\prod_{j\in[\![p]\!]} n_j} + C_2 \sum_{j\in[\![p]\!]} \frac{1}{n_j},
\end{aligned}$$

with $C_1 = \frac{8}{(2\pi)^d}$ and $C_2 = 4p3^{2(p-2)}(2\pi)^d \max_{j\in[\![p]\!]} \|f_j\|_{\mathbb{L}^2_d}^2 + \frac{C_0 2d^2 6^{d-1}\kappa_\varphi^2}{\xi^2}$, which proves the result.

□

Remark 2.2.7:
Analogously to the discussion of convergence rates in Section 2.1.3, it is also possible to consider d-dimensional Hermite-Sobolev spaces as an extension of such one-dimensional spaces defined and studied in Bongioanni and Torrea (2006), introducing a condition on the coefficient leading to

$$\|g^{(p)}_{\boldsymbol{m}} - g^{(p)}\|^2_{\mathbb{L}^2_d} \leq C(m_1^{-\alpha_1} + \ldots + m_d^{-\alpha_d}).$$

As for Proposition 2.2.5, choosing $\boldsymbol{m}_{\text{opt}}$ via $\sqrt{m_{k,\text{opt}}} = h^{-1}_{k,\text{opt}}$ for $k \in [\![d]\!]$, we get the same rates as in Section 2.1.3. However, the link between the behavior of the coefficients and functional regularity is not proven in the multivariate setting.

2.2.4 Model selection

Since the optimal parameter for the dimension $\boldsymbol{m}_{\text{opt}}$, see Remark 2.2.7, depends on the regularity of $g^{(p)}$, which is in general unknown, we consider a method to select the parameter only based on data in this section. For simplicity, we consider the case where all sample sizes are the same, i.e., $n_1 = \ldots = n_p = n \in \mathbb{N}$. The results of this section can be extended to the case that sample sizes are of the same order leading only to changes in the constants. In addition, we keep using the Hermite basis $(\phi_{\boldsymbol{\ell}})_{\boldsymbol{\ell}\geq 0}$ of $\mathbb{L}^2_d$ and the estimator $\widehat{g^{(p)}_{\boldsymbol{m}}}$ given by (2.35). Let us motivate the estimation approach. The idea is to look for a value of $\boldsymbol{m} \in \mathbb{N}^d$ such that

$$\|g^{(p)} - \widehat{g^{(p)}_{\boldsymbol{m}}}\|^2_{\mathbb{L}^2_d} = \|g^{(p)} - g^{(p)}_{\boldsymbol{m}}\|^2_{\mathbb{L}^2_d} + \|g^{(p)}_{\boldsymbol{m}} - \widehat{g^{(p)}_{\boldsymbol{m}}}\|^2_{\mathbb{L}^2_d} \tag{2.44}$$

is as small as possible. Note that

$$\langle g^{(p)}, g^{(p)}_{\boldsymbol{m}}\rangle_{\mathbb{L}^2_d} = \sum_{\boldsymbol{\ell}_1=0}^{\boldsymbol{m}-1} \sum_{\boldsymbol{\ell}_2\geq 0} a_{\boldsymbol{\ell}_1} a_{\boldsymbol{\ell}_2} \int_{\mathbb{R}^d} \phi_{\boldsymbol{\ell}_1}(\boldsymbol{t})\phi_{\boldsymbol{\ell}_2}(\boldsymbol{t}) d\lambda^d(\boldsymbol{t}) = \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} |a_{\boldsymbol{\ell}}|^2 = \|g^{(p)}_{\boldsymbol{m}}\|^2.$$

Consequently, it holds that

$$\|g^{(p)} - g^{(p)}_{\boldsymbol{m}}\|^2_{\mathbb{L}^2_d} = \|g^{(p)}\|^2_{\mathbb{L}^2_d} - 2\langle g^{(p)}_{\boldsymbol{m}}, g^{(p)}\rangle_{\mathbb{L}^2_d} + \|g^{(p)}_{\boldsymbol{m}}\|^2_{\mathbb{L}^2_d} = \|g^{(p)}\|^2_{\mathbb{L}^2_d} - \|g^{(p)}_{\boldsymbol{m}}\|^2_{\mathbb{L}^2_d}.$$

and minimizing $\|g^{(p)} - \widehat{g^{(p)}_{\boldsymbol{m}}}\|^2_{\mathbb{L}^2_d}$ over $\boldsymbol{m}$ is equivalent to minimizing

$$\|g^{(p)}_{\boldsymbol{m}} - \widehat{g^{(p)}_{\boldsymbol{m}}}\|^2_{\mathbb{L}^2_d} - \|g^{(p)}_{\boldsymbol{m}}\|^2_{\mathbb{L}^2_d}. \tag{2.45}$$

To estimate the first term, we use the order of its expectation derived in Proposition 2.2.5. This term we call penalty and denote the sequence $(\text{pen}(\boldsymbol{m}))_{\boldsymbol{m}\in\mathbb{N}^d}$. It is given for some $\kappa \in \mathbb{R}_+$ by

$$\text{pen}(\boldsymbol{m}) = \kappa \frac{\prod_{k\in[\![d]\!]} \sqrt{2m_k + 1} \log^d(n)}{n^d}.$$

Note that in comparison to the bound for the variance in Proposition 2.2.5 an additional factor $\log^d(n)$ is needed, as can be seen in the proof. Further, the factor κ is a numerical constant which does not depend on the data and is given in the next result. However, in practice smaller values work better. The value κ is calibrated once for all, through intensive simulation experiments or with specific methods (Comte (2017) and Baudry et al. (2012)). In our small simulation experiments, see Section 2.3 for more details, $\kappa = 3$ seems to work. The theoretical value κ_0 is likely to be too large in practice.

For the second term in Equation (2.45), recall that $\widehat{g_{\boldsymbol{m}}^{(p)}}$ minimizes the contrast given for $t \in S_{\boldsymbol{m}}$ by

$$\gamma_n(t) := \|t\|_{\mathbb{L}^2_d}^2 - \frac{2}{n^p} \sum_{k_1,\dots,k_p \in [\![n]\!]} t(\boldsymbol{X}_{k_1}^{(1)} + \ldots + \boldsymbol{X}_{k_p}^{(p)}).$$

Consequently, we estimate $-\|g_{\boldsymbol{m}}^{(p)}\|_{\mathbb{L}^2_d}^2$ by $-\|\widehat{g_{\boldsymbol{m}}^{(p)}}\|_{\mathbb{L}^2_d}^2 = \gamma_n(\widehat{g_{\boldsymbol{m}}^{(p)}}) = \min_{t \in S_{\boldsymbol{m}}} \gamma_n(t)$. Finally, define

$$\widehat{\boldsymbol{m}} = \arg\min_{\boldsymbol{m} \in \mathcal{M}_n} \{\gamma_n(\widehat{g_{\boldsymbol{m}}^{(p)}}) + \mathrm{pen}(\boldsymbol{m})\} \tag{2.46}$$

where $\mathcal{M}_n \subset \mathbb{N}^d$. We take $\mathcal{M}_n$ as the collection of models such that, for $\mathfrak{c}_0 := (d^2 - p + 1)/(2\xi)$,

$$\mathcal{M}_n := \left\{ \boldsymbol{m} \in \mathbb{N}^d : \forall k \in [\![d]\!], m_k \geq \mathfrak{c}_0 \log(n) \text{ and } \prod_{k \in [\![d]\!]} \sqrt{2m_k + 1} \leq n^d \right\}. \tag{2.47}$$

Note the similarity of the collection to the assumption made on the dimension parameter in Equation (2.37) and see the comments herewith. It is slightly reinforced taking in account that the sample sizes are assumed to be equal in this section.

Theorem 2.2.8:
Assume that $\mathcal{F}[f_j] \in \mathbb{L}^1_d \cap \mathbb{L}^2_d$ for all $j \in [\![p]\!]$, and consider the estimator $\widehat{g_{\widehat{\boldsymbol{m}}}^{(p)}}$ of $g^{(p)}$ defined by (2.35)-(2.46) with $\phi_{\boldsymbol{\ell}}$ as in (2.28)-(2.33). Then, there exists κ_0 such that for any $\kappa \geq \kappa_0$,

$$\mathbb{E}\left[\|\widehat{g_{\widehat{\boldsymbol{m}}}^{(p)}} - g^{(p)}\|_{\mathbb{L}^2_d}^2\right] \leq \inf_{\boldsymbol{m} \in \mathcal{M}_n} \left(9\mathbb{E}[\|\widehat{g_{\boldsymbol{m}}^{(p)}} - g^{(p)}\|_{\mathbb{L}^2_d}^2] + 2\mathrm{pen}(\boldsymbol{m})\right) + \frac{C}{n},$$

where $\kappa_0 = 11\frac{2^d(4(pd-1))^p}{\pi^d}$ suits, and C is a constant depending on $p, d, \xi, \kappa_\varphi$ and $\|f_j\|_{\mathbb{L}^2_d}$, $j \in [\![p]\!]$.

Remark 2.2.9:
Proposition 2.2.5 gives an upper bound for $\mathbb{E}[\|\widehat{g_{\boldsymbol{m}}^{(p)}} - g^{(p)}\|_{\mathbb{L}^2_d}^2]$. Therefore, Theorem 2.2.8 shows that, up to a multiplicative constant, the risk of the final estimator minimizes over all $\boldsymbol{m}$'s in the collection, the squared-bias $\|g_{\boldsymbol{m}}^{(p)} - g^{(p)}\|_{\mathbb{L}^2_d}^2$ plus the penalty which has the order of the variance up to a $\log^d(n)$ term. However, this log-loss does not change the rates.

Proof of Theorem 2.2.8. We consider a nesting space, i.e., a space $S_{\boldsymbol{M}}$ such that, for all $\boldsymbol{m} \in \mathcal{M}_n$, $S_{\boldsymbol{m}} \subset S_{\boldsymbol{M}}$, $\boldsymbol{M} = (M_1, \ldots, M_d)$. As our worst case assumption for the spaces $S_{\boldsymbol{m}}$ is $\sqrt{2m_j + 1} \leq n^d$ in case $\boldsymbol{m} = (m_1, \ldots, m_d) \in \mathcal{M}_n$ is such that all m_k are equal to 1 except m_j, we set $M_j = \lfloor (n^{2d} - 1)/2 \rfloor$ for all $j \in [\![p]\!]$, where $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$. The remaining part of the proof is structured in five steps, first splitting the expectation to bound in multiple summands

and then upper bounding them separately.
Step 1. It holds for all $\boldsymbol{m} \in \mathcal{M}_n$ that

$$\mathbb{E}\left[\|\widehat{g^{(p)}_{\widehat{\boldsymbol{m}}}} - g^{(p)}\|^2_{\mathbb{L}^2_d}\right] \leq 3\mathbb{E}\left[\|\widehat{g^{(p)}_{\boldsymbol{m}}} - g^{(p)}\|^2_{\mathbb{L}^2_d}\right] + 6\mathbb{E}\left[\|\widehat{g^{(p)}_{\boldsymbol{m}}} - g^{(p)}_{\boldsymbol{m}}\|^2_{\mathbb{L}^2_d}\right] + 2\text{pen}(\boldsymbol{m})$$
$$+11\mathbb{E}\left[\left(\|\widehat{g^{(p)}_{\widehat{\boldsymbol{m}}}} - g^{(p)}_{\widehat{\boldsymbol{m}}}\|^2_{\mathbb{L}^2_d} - \frac{2}{11}\text{pen}(\widehat{\boldsymbol{m}})\right)_+\right].$$

See the proof of Theorem 2 in Brenner Miguel (2022) for detailed calculations. The proof relies on the fact that there exits a nesting space $S_{\boldsymbol{M}}$ and uses resulting norm inequalities. Let us set

$$\mathbb{E}_0 := \mathbb{E}\left[\left(\|\widehat{g_{\widehat{\boldsymbol{m}}}}^{(p)} - g^{(p)}_{\widehat{\boldsymbol{m}}}\|^2_{\mathbb{L}^2_d} - \frac{2}{11}\text{pen}(\widehat{\boldsymbol{m}})\right)_+\right],$$

which is the term that remains to be bounded. We will show $\mathbb{E}_0 \lesssim n^{-1}$ which then yields the result.

Step 2. Decomposition of $\mathbb{E}_0$. It holds that

$$\|\widehat{g^{(p)}_{\boldsymbol{m}}} - g^{(p)}_{\boldsymbol{m}}\|^2_{\mathbb{L}^2_d} = \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1}(\widehat{a}_{\boldsymbol{\ell}} - a_{\boldsymbol{\ell}})^2 = \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1}\nu_n^2(\phi_{\boldsymbol{\ell}})$$

where

$$\nu_n(t) := \frac{1}{n^p}\sum_{k\in[\![p]\!],j_k\in[\![n]\!]}\left(t(\boldsymbol{X}^{(1)}_{j_1} + \cdots + \boldsymbol{X}^{(p)}_{j_p}) - \langle t, g^{(p)}\rangle_{\mathbb{L}^2_d}\right).$$

Note, that for any $t \in S_{\boldsymbol{m}}$ it holds that there exist coefficients $b_{\boldsymbol{\ell}}$ such that $t = \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} b_{\boldsymbol{\ell}}\phi_{\boldsymbol{\ell}}$. Further, if $\|t\|^2_{\mathbb{L}^2_d} = 1$, if follows $\sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} b^2_{\boldsymbol{\ell}} = 1$ and, consequently, with Cauchy–Schwarz inequality we get

$$\begin{aligned}\nu_n^2(t) &= \left(\frac{1}{n^p}\sum_{k\in[\![p]\!],j_k\in[\![n]\!]}\left(\sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} b_{\boldsymbol{\ell}}\phi_{\boldsymbol{\ell}}(\boldsymbol{X}^{(1)}_{j_1} + \cdots + \boldsymbol{X}^{(p)}_{j_p}) - \langle\sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} b_{\boldsymbol{\ell}}\phi_{\boldsymbol{\ell}}, g^{(p)}\rangle_{\mathbb{L}^2_d}\right)\right)^2 \\ &= \left(\sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} b_{\boldsymbol{\ell}}\frac{1}{n^p}\sum_{k\in[\![p]\!],j_k\in[\![n]\!]}\left(\phi_{\boldsymbol{\ell}}(\boldsymbol{X}^{(1)}_{j_1} + \cdots + \boldsymbol{X}^{(p)}_{j_p}) - \langle\phi_{\boldsymbol{\ell}}, g^{(p)}\rangle_{\mathbb{L}^2_d}\right)\right)^2 \\ &= \left(\sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} b_{\boldsymbol{\ell}}\nu_n(\phi_{\boldsymbol{\ell}})\right)^2 \leq \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1} b^2_{\boldsymbol{\ell}}\sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1}\nu_n^2(\varphi_\ell) = \sum_{\boldsymbol{\ell}=0}^{\boldsymbol{m}-1}\nu_n^2(\varphi_\ell).\end{aligned}$$

Therefore we need to bound

$$\mathbb{E}_0 = \mathbb{E}\left[\left(\sup_{t\in S_{\widehat{\boldsymbol{m}}},\|t\|_{\mathbb{L}^2_d}=1}\nu_n^2(t) - \frac{2}{11}\text{pen}(\widehat{\boldsymbol{m}})\right)_+\right].$$

We rewrite $\nu_n(t)$, namely,

$$\begin{aligned}\nu_n(t) &= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\left(\frac{1}{n^p}\sum_{k\in[\![p]\!],j_k\in[\![n]\!]} e^{i\langle\boldsymbol{X}^{(1)}_{j_1}+\cdots+\boldsymbol{X}^{(p)}_{j_p},\boldsymbol{x}\rangle} - \mathcal{F}[g^{(p)}](\boldsymbol{x})\right)\mathcal{F}[t](-\boldsymbol{x})d\lambda^d(\boldsymbol{x}) \\ &= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\left[\prod_{j\in[\![p]\!]}\widehat{\mathcal{F}[f_j]}(\boldsymbol{x}) - \prod_{j\in[\![p]\!]}\mathcal{F}[f_j](\boldsymbol{x})\right]\mathcal{F}[t](-\boldsymbol{x})d\lambda^d(\boldsymbol{x}).\end{aligned}$$

We define for some $\tau \in \mathbb{R}_+$, which will be chosen later the set

$$\Omega(\boldsymbol{x}) := \bigcap_{j \in [\![p]\!]} \Omega_j(\boldsymbol{x}), \quad \Omega_j(\boldsymbol{x}) = \left\{ |\widehat{\mathcal{F}[f_j]}(\boldsymbol{x}) - \mathcal{F}[f_j](\boldsymbol{x})| \leq \tau \sqrt{\frac{\log(n)}{n}} \right\}. \tag{2.48}$$

Then we consider the decomposition (where $\boldsymbol{u} \leq \boldsymbol{v}$ is to be understood component-wise if $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors)

$$\nu_n(t) = \nu_{n,1}(t) + \nu_{n,2}(t) + \nu_{n,3}(t) + \nu_{n,4}(t)$$

with

$$\nu_{n,1}(t) := \frac{1}{(2\pi)^d} \int_{|\boldsymbol{x}| \leq \sqrt{2\widehat{\boldsymbol{m}}+1}} \prod_{j \in [\![p]\!]} \left[\widehat{\mathcal{F}[f_j]}(\boldsymbol{x}) - \mathcal{F}[f_j](\boldsymbol{x}) \right] \mathcal{F}[t](-\boldsymbol{x}) \mathbb{1}_{\Omega(\boldsymbol{x})} d\lambda^d(\boldsymbol{x}),$$

$$\nu_{n,2}(t) := \frac{1}{(2\pi)^d} \int_{|\boldsymbol{x}| \leq \sqrt{2\widehat{\boldsymbol{m}}+1}} \prod_{j \in [\![p]\!]} \left[\widehat{\mathcal{F}[f_j]}(\boldsymbol{x}) - \mathcal{F}[f_j](\boldsymbol{x}) \right] \mathcal{F}[t](-\boldsymbol{x}) \mathbb{1}_{\Omega(\boldsymbol{x})^c} d\lambda^d(\boldsymbol{x}),$$

$$\nu_{n,3}(t) := \frac{1}{(2\pi)^d} \int_{\exists j, |x_j| > \sqrt{2\widehat{m}_j+1}} \prod_{j \in [\![p]\!]} \left[\widehat{\mathcal{F}[f_j]}(\boldsymbol{x}) - \mathcal{F}[f_j](\boldsymbol{x}) \right] \mathcal{F}[t](-\boldsymbol{x}) d\lambda^d(\boldsymbol{x})$$

and

$$\begin{aligned}
\nu_{n,4}(t) &:= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} (\widehat{\Delta\mathcal{F}[f]})_2(\boldsymbol{x}) \mathcal{F}[t](-\boldsymbol{x}) d\lambda^d(\boldsymbol{x}), \\
(\widehat{\Delta\mathcal{F}[f]})_2 &:= \left[\prod_{j \in [\![p]\!]} \widehat{\mathcal{F}[f_j]} - \prod_{j \in [\![p]\!]} \mathcal{F}[f_j] \right] - \prod_{j \in [\![p]\!]} \left[\widehat{\mathcal{F}[f_j]} - \mathcal{F}[f_j] \right].
\end{aligned}$$

We get

$$\begin{aligned}
\mathbb{E}_0 &\leq 2\mathbb{E}\left[\left(\sup_{t \in S_{\widehat{\boldsymbol{m}}}, \|t\|_{\mathbb{L}^2_d}=1} \nu_{n,1}^2(t) - \frac{1}{11} \operatorname{pen}(\widehat{\boldsymbol{m}}) \right)_+ \right] + 6 \sum_{k=2}^{4} \mathbb{E}\left[\sup_{t \in S_{\widehat{\boldsymbol{m}}}, \|t\|_{\mathbb{L}^2_d}=1} \nu_{n,k}^2(t) \right] \\
&=: 2\mathbb{E}_1 + 6(\mathbb{E}_2 + \mathbb{E}_3 + \mathbb{E}_4),
\end{aligned}$$

and we successively study the four terms.

Step 3. The bound on the main term $\mathbb{E}_1$ is straightforward. Indeed, using first the Cauchy–

Schwarz inequality, and then taking the definition of $\Omega(\boldsymbol{x})$ into account, we get

$$\sup_{t\in S_{\widehat{m}},\|t\|_{\mathbb{L}^2_d}=1} \nu_{n,1}^2(t) \leq \sup_{t\in S_{\widehat{\boldsymbol{m}}},\|t\|_{\mathbb{L}^2_d}=1} \frac{1}{(2\pi)^{2d}} \int_{\mathbb{R}^d} |\mathcal{F}[t](\boldsymbol{x})|^2 d\lambda^d(\boldsymbol{x})$$

$$\int_{|\boldsymbol{x}|\leq\sqrt{2\widehat{\boldsymbol{m}}+1}} \left| \prod_{j\in[\![p]\!]} \left(\widehat{\mathcal{F}[f_j]}(\boldsymbol{x}) - \mathcal{F}[f_j](\boldsymbol{x})\right) \right|^2 \mathbb{1}_{\Omega(\boldsymbol{x})} d\lambda^d(\boldsymbol{x})$$

$$\leq \frac{1}{(2\pi)^d} \int_{|\boldsymbol{x}|\leq\sqrt{2\widehat{\boldsymbol{m}}1}} \left| \prod_{j\in[\![p]\!]} \tau\sqrt{\frac{\log(n)}{n}} \right|^2 d\lambda^d(\boldsymbol{x})$$

$$= \frac{\tau^{2p}}{(2\pi)^d} \prod_{k\in[\![d]\!]} 2\sqrt{2\widehat{m}_k+1} \frac{\log^p(n)}{n^p}$$

Now, if κ is such that the above quantity is less than $(1/11)\,\mathrm{pen}(\widehat{\boldsymbol{m}})$, that is,

$$\frac{2^d\tau^{2p}}{(2\pi)^d} \leq \kappa/11, \quad \text{i.e.,} \quad \kappa \geq \kappa_0 := 11\frac{2^d\tau^{2p}}{(2\pi)^d}$$

we obtain that $\mathbb{E}_1 \leq 0$ and therefore $\mathbb{E}_1 = 0$.

Step 4. The bound on $\mathbb{E}_2$ relies on Lemma 1 p.12 in Ammous et al. (2024), which states that for any $\tau \in \mathbb{R}_+$, $\boldsymbol{x} \in \mathbb{R}^d$, $d \geq 1$,

$$\mathbb{P}(\Omega_j(\boldsymbol{x})^c) \leq 4n^{-\tau^2/4}.$$

Therefore, $\mathbb{P}(\Omega(\boldsymbol{x})^c) \leq 4pn^{-\tau^2/4}$. Due to the boundedness of the Fourier transform by one, note that

$$\left| \prod_{j\in[\![p]\!]} \left[\widehat{\mathcal{F}[f_j]}(\boldsymbol{x}) - \mathcal{F}[f_j](\boldsymbol{x})\right] \right|^2 \leq 2^{2p}.$$

Additionally, applying Cauchy–Schwarz inequality, for $t \in S_{\widehat{\boldsymbol{m}}}$ with $\|t\|_{\mathbb{L}^2_d} = (2\pi)^{-d}\|\mathcal{F}[t]\|_{\mathbb{L}^2_d} = 1$ we get that

$$\nu_{n,2}^2(t) = \left(\frac{1}{(2\pi)^d} \int_{|\boldsymbol{x}|\leq\sqrt{2\widehat{\boldsymbol{m}}+1}} \prod_{j\in[\![p]\!]} \left[\widehat{\mathcal{F}[f_j]}(\boldsymbol{x}) - \mathcal{F}[f_j](\boldsymbol{x})\right] \mathcal{F}[t](-\boldsymbol{x}) \mathbb{1}_{\Omega(\boldsymbol{x})^c} d\lambda^d(\boldsymbol{x}) \right)^2$$

$$\leq \left(\frac{1}{(2\pi)^d} \int_{|\boldsymbol{x}|\leq\sqrt{2\widehat{\boldsymbol{m}}+1}} \left| \prod_{j\in[\![p]\!]} \left[\widehat{\mathcal{F}[f_j]}(\boldsymbol{x}) - \mathcal{F}[f_j](\boldsymbol{x})\right] \right|^2 \mathbb{1}_{\Omega(\boldsymbol{x})^c} d\lambda^d(\boldsymbol{x}) \right)$$

$$\cdot \left(\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\mathcal{F}[t](-\boldsymbol{x})|^2 d\lambda^d(\boldsymbol{x}) \right)$$

$$\leq \frac{1}{(2\pi)^d} \int_{|\boldsymbol{x}|\leq\sqrt{2\widehat{\boldsymbol{m}}+1}} 2^{2p} \mathbb{1}_{\Omega(\boldsymbol{x})^c} d\lambda^d(\boldsymbol{x})$$

and

$$\mathbb{E}_2 \leq \frac{2^{2p}}{(2\pi)^d} \int_{|\boldsymbol{x}|\leq\sqrt{2\boldsymbol{M}+1}} \mathbb{P}\left(\Omega(\boldsymbol{x})^c\right) d\lambda^d(\boldsymbol{x}) \leq 4\frac{p2^{2p+2d}}{(2\pi)^d} \prod_{k\in[\![d]\!]} \sqrt{2M_k+1} n^{-\tau^2/4}.$$

Using that per definition $\sqrt{2M_k+1} \leq n^d$ for all $k \in [\![d]\!]$, we have $\mathbb{E}_2 \lesssim n^{d^2-\tau^2/4}$. Thus $\mathbb{E}_2 \lesssim n^{-1}$ for

$$\tau^2 \geq 4(d^2-1).$$

Step 5. The study of $\mathbb{E}_3$ follows the line of the study of $T_{2,1}$ in the proof of Proposition 2.2.5. For $t \in S_{\widehat{\boldsymbol{m}}}$, $\|t\|_{\mathbb{L}^2_d} = 1$ writing $t = \sum_{\boldsymbol{\ell}=0}^{\widehat{\boldsymbol{m}}-1} a_{\boldsymbol{\ell}}\phi_{\boldsymbol{\ell}}$ it follows for its Fourier transform that

$$\mathcal{F}[t] = \sum_{\boldsymbol{\ell}=0}^{\widehat{\boldsymbol{m}}-1} a_{\boldsymbol{\ell}}\mathcal{F}[\phi_{\boldsymbol{\ell}}].$$

Additionally, applying Cauchy–Schwarz inequality twice, we get that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in S_{\widehat{\boldsymbol{m}}},\|t\|_{\mathbb{L}^2_d}=1} \nu_{n,3}^2(t)\right] \\
&= \frac{1}{(2\pi)^d}\mathbb{E}\left[\sup_{t\in S_{\widehat{\boldsymbol{m}}},\|t\|_{\mathbb{L}^2_d}=1} \left(\sum_{\boldsymbol{\ell}=0}^{\widehat{\boldsymbol{m}}-1} |a_{\boldsymbol{\ell}}|^2\right)\right. \\
&\qquad\left.\cdot\left(\sum_{\boldsymbol{\ell}=0}^{\widehat{\boldsymbol{m}}-1}\left|\int_{\exists i,|x_i|>\sqrt{2\widehat{m}_i+1}} \prod_{j\in[\![p]\!]}\left[\widehat{\mathcal{F}[f_j]}(\boldsymbol{x}) - \mathcal{F}[f_j](\boldsymbol{x})\right]\phi_{\boldsymbol{\ell}}(\boldsymbol{x})d\lambda^d(\boldsymbol{x})\right|^2\right)\right],
\end{aligned}$$

and using also that $\|t\|^2_{\mathbb{L}^2_d} = \sum_{\boldsymbol{\ell}=0}^{\widehat{\boldsymbol{m}}-1} |a_{\boldsymbol{\ell}}|^2 = 1$ it follows

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in S_{\widehat{\boldsymbol{m}}},\|t\|_{\mathbb{L}^2_d}=1} \nu_{n,3}^2(t)\right] \\
&\leq 2d\sum_{k\in[\![d]\!]}\mathbb{E}\left[\sum_{\boldsymbol{\ell}=0}^{\widehat{\boldsymbol{m}}-1}\int_{\mathbb{R}^d}\mathbb{1}_{\{|x_k|>\sqrt{2\widehat{m}_k+1}\}}\left|\left(\widehat{\Delta\mathcal{F}[f]}\right)_1(\boldsymbol{x})\right|^2|\phi_{\boldsymbol{\ell}}(\boldsymbol{x})|d\lambda^d(\boldsymbol{x})\right. \\
&\qquad\left.\cdot\int_{\mathbb{R}^d}\mathbb{1}_{\{|x_k|>\sqrt{2\widehat{m}_k+1}\}}|\phi_{\boldsymbol{\ell}}(\boldsymbol{x})|d\lambda^d(\boldsymbol{x})\right].
\end{aligned}$$

Now we use that $\widehat{m}_k \geq \mathfrak{c}_0\log(n)$ and $\prod_{k\in[\![d]\!]}\sqrt{2\widehat{m}_k+1} \leq n^d$ as $\widehat{\boldsymbol{m}} \in \mathcal{M}_n$ given by (2.47) and we get with κ_φ defined in (2.30)

$$\mathbb{E}_3 \leq \frac{2d^2 6^{2(d-1)}\kappa_\varphi^2}{\xi^2}\frac{\prod_{k\in[\![d]\!]}\sqrt{2M_k+1}}{n^p}\sum_{k\in[\![d]\!]}\exp\left(-2\xi\mathfrak{c}_0\log(n)\right) = \frac{2d^2 6^{2(d-1)}d\kappa_\varphi^2}{\xi^2}\frac{1}{n},$$

due to the definition of $\mathfrak{c}_0$. This ensures that $\mathbb{E}_3 \lesssim 1/n$.

Step 6. The study of $\mathbb{E}_4$ is also rather straightforward. Indeed, by Cauchy–Schwarz Inequality

$$\mathbb{E}_4 \leq \mathbb{E}\left[\sup_{t\in S_{\widehat{\boldsymbol{m}}}, \|t\|_{\mathbb{L}^2_d}=1} \frac{1}{(2\pi)^{2d}} \int_{\mathbb{R}^d} |\mathcal{F}[t](\boldsymbol{x})|^2 d\lambda^d(\boldsymbol{x}) \int_{\mathbb{R}^d} (\widehat{\Delta\mathcal{F}[f]})_2^2(\boldsymbol{x}) d\lambda^d(\boldsymbol{x})\right]$$
$$\leq \frac{1}{(2\pi)^d}\mathbb{E}\left[\int_{\mathbb{R}^d} (\widehat{\Delta\mathcal{F}[f]})_2^2(\boldsymbol{x}) d\lambda^d(\boldsymbol{x})\right].$$

Now using Remark 2.4.2 (revisited), we get

$$\mathbb{E}_4 \leq \frac{p3^{2(p-2)}}{(2\pi)^d} \sum_{j\in[\![p]\!]} \frac{1}{n} \int_{\mathbb{R}^d} |\mathcal{F}[f_{j+1}](\boldsymbol{x})|^2 d\lambda^d(\boldsymbol{x}) \leq p^2 3^{2(p-2)} \max_{j\in[\![p]\!]}(\|f_j\|^2_{\mathbb{L}^2_d}) \frac{1}{n},$$

as we assumed that all n_j are equal to n. This concludes the proof.

□

2.3 Simulation study

We illustrate the results of Section 2.1 and Section 2.2 in a Monte Carlo simulation study. Our simulation experiments deal with the cases where $d = 1$ and $p = 2$ or $p = 3$. First, we consider examples for which we analytically derive the convolution to be able to evaluate the procedures.

(i) Two independent Gaussian distributed random variables.

(ii) Two independent Claw distributed random variables. The Claw distribution is defined as a specific sum of Gaussian random variables. More precisely, for a standard Gaussian distributed random variable X and Gaussian random variables $Z_l \sim \mathcal{N}(l/2 - 1, 1/4)$ for $l \in \mathbb{N}$, a Claw distributed random variable W satisfies $W = 0.5X + \sum_{l=0}^4 Z_l$ in distribution.

(iii) Two independent random variables uniformly distributed on the interval $[-1, 1]$, denoted by $\mathrm{Unif}[-1, 1]$.

(iv) Three independent $\mathrm{Unif}[-1, 1]$ distributed random variables.

Examples (i)-(iii) are the examples given in Chesneau et al. (2013). For all cases we consider the Kernel estimator defined in Equation (2.11) and the projection estimator defined in Equation (2.35). As kernel we consider a kernel of order 5 given by $L(x) = 3\mathrm{n}_1(x) - 3\mathrm{n}_2(x) + \mathrm{n}_3(x)$. Here, n_j denotes the density of a centered Gaussian random variable with variance $j \in \mathbb{N}$. Note that, similarly, one can build kernels of higher order; see Comte and Marie (2020) for details. For the projection method, we use the Hermite basis throughout the simulations.

Figure 2.1 below shows the behavior of kernel estimators. In the left column of Figure 2.1, kernel estimators corresponding to the bandwidth $h \in \mathcal{H}$ are depicted, where $\mathcal{H}$ denotes the set of 20 equidistant points starting from 0.005 and ending with 1.25. It visualizes the necessity of a data-driven choice of the parameter h to obtain a trade-off between over-fitting (dark green lines) and a too rough estimator (light green lines). In the right column of Figure 2.1, the curve represents the MISE computed over 50 Monte Carlo simulations as a function of h. One can see that in the given examples, there seem to be intervals of appropriate choices for h of different sizes depending

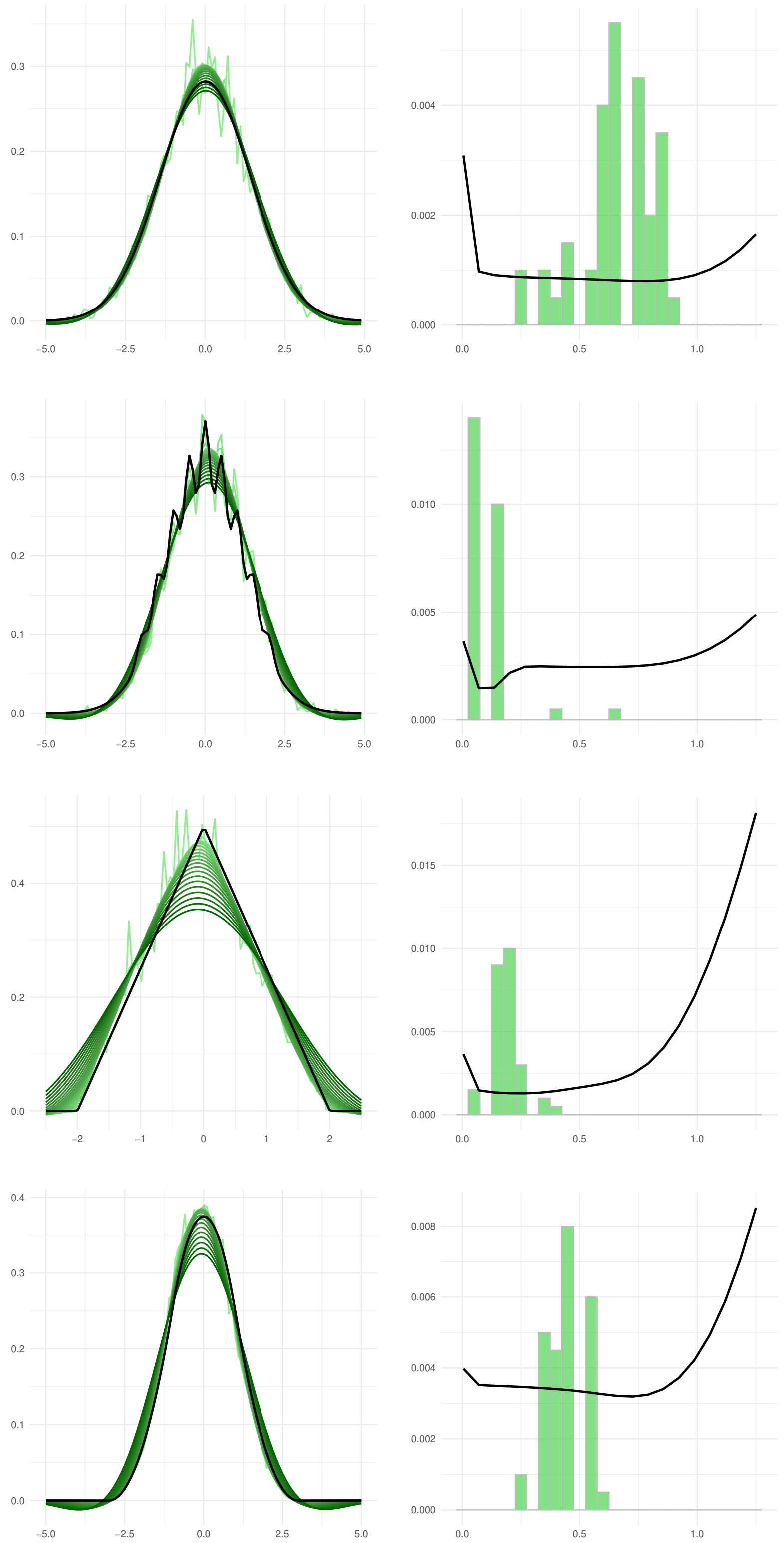


Figure 2.1: Kernel estimator with $n = 200$ for examples (i)-(iii) and $n = 60$ for (iv); Left column: Plot of $\widehat{g}_h$ for $h \in \mathcal{H}$ (green lines) and true function g (black line); Right column: Estimated MISE (black line) using $M = 50$ Monte Carlo simulations and estimated values $\widehat{h}$ using the PCO-method (green histograms).

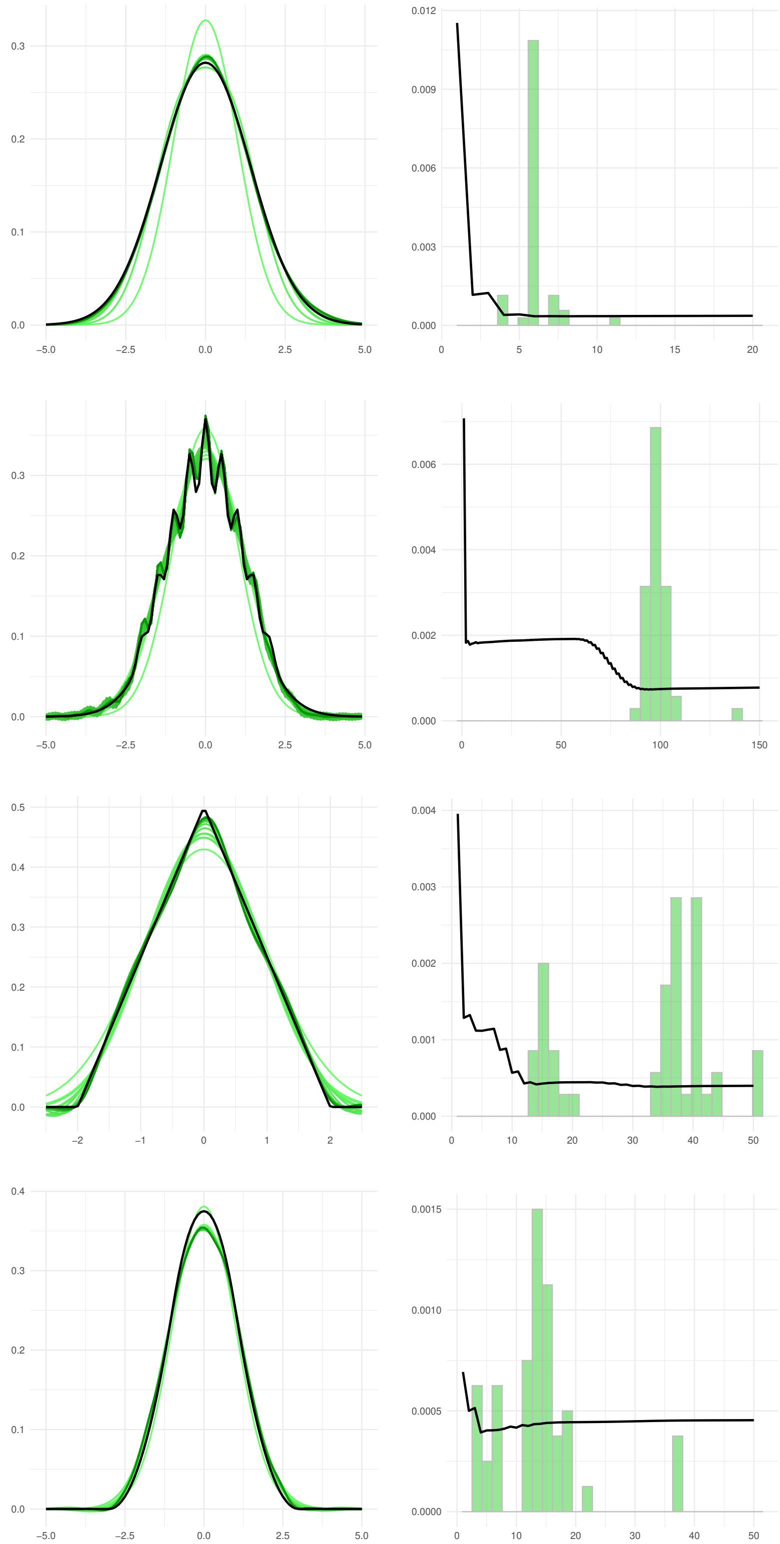


Figure 2.2: Projection estimator with $n = 200$ for examples (i)–(iii) and $n = 60$ for (iv); Left column: Plot of $\widehat{g_m}$ for $m \in [\![20]\!]$, $[\![50]\!]$ and $[\![150]\!]$, respectively (green lines) and true function g (black line); Right column: Estimated MISE plot with $M = 50$ Monte Carlo simulations and estimated values $\widehat{m}$ using the selection method (green histograms).

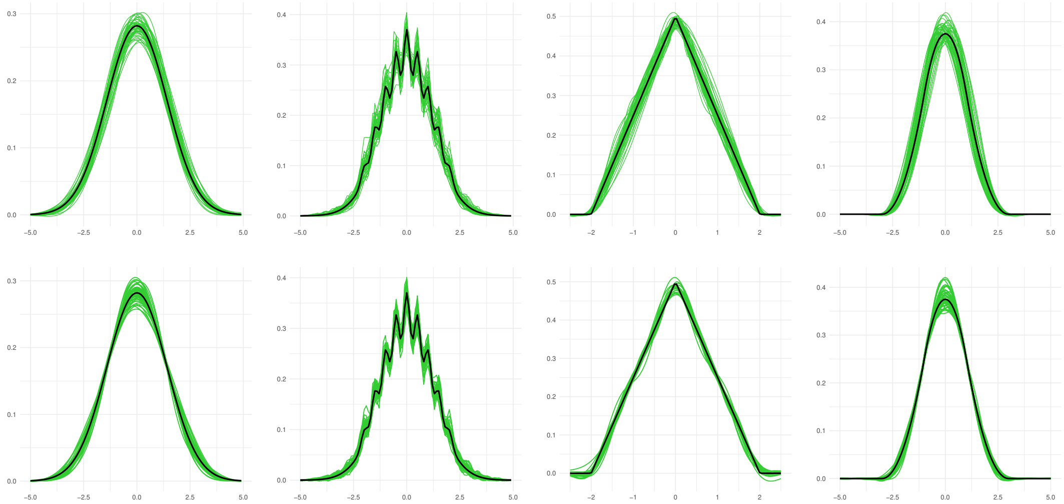


Figure 2.3: Plots of $\widehat{g_{\widehat{h}}}$ using the PCO-method (green lines, first row) and $\widehat{g_{\widehat{m}}}$ using selection method (second line) for $M = 50$ Monte Carlo simulations and $n = 200$ for Examples (i)–(iii) and $n = 60$ for (iv).

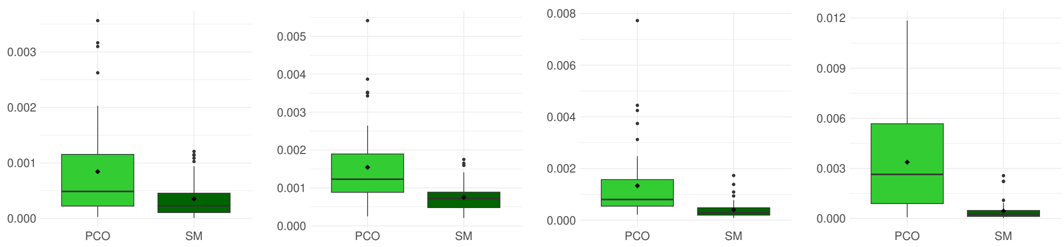


Figure 2.4: Estimated integrated squared error choosing h with PCO-Method for the kernel estimator and m with the selection method for the projection estimator with Hermite basis for $M = 50$ Monte Carlo simulations; $n = 200$ for examples (i)–(iii) and $n = 60$ for example (iv).

on the example; furthermore, the theory suggests choosing the smallest bandwidth $h = 1/n$, but this choice is revealed to be too small in these examples. For instance, in the first three examples this corresponds to 0.005 which is the first element of the considered interval, i.e., the first point of each black line in the plots. The histograms depict the choices for h made with the PCO-method described in Section 2.1.5. They show that the data-driven method performs quite well on these examples. Values of bandwidth outside the optimal interval occur only for the Claw-distribution example, and only for two out of 50 Monte Carlo iterations.

Figure 2.2 depicts similar results for the projection estimator and the model selection method proposed in Section 2.2.4. Here, apart from finding a sufficiently good estimator, the goal is to find a parsimonious decomposition of $g^{(p)}$, i.e., a small number m of coefficients for estimating the function $g^{(p)}$ appropriately well. For examples (i), (ii) and (iv), this goal seems to be achieved.

Figure 2.3 and Figure 2.4 compare the performance of the projection and the kernel method. Slightly better results over all examples and the significantly shorter computation time motivated the focus of the theoretical consideration on the selection method in Section 2.2.4.

	(i) Gauss + Gauss			(ii) Claw + Claw			(iii) Unif + Unif		
n	K	P	A	K	P	A	K	P	A
50	4.04	0.18	0.05	4.15	0.36	0.27	4.56	0.24	2.01
100	0.45	0.07	0.01	0.48	0.23	0.08	0.62	0.15	1.98
200	0.08	0.04	0.00	0.15	0.08	0.02	0.13	0.05	1.98

	Gam + Gam			Exp + Exp		
n	K	P	A	K	P	A
50	10.37	2.35	4.92	4.77	6.53	0.24
100	6.21	1.82	4.78	0.83	5.85	0.10
200	5.53	1.54	4.72	0.21	5.68	0.06

Table 2.1: Errors for Examples (i)–(iii), a Gamma$(2, 1)$ and a Gamma$(1, 1)$ distributed random variable (Gam + Gam) $\mathcal{T} = [0, 20]$ and two Exp(1) distributed random variables (Exp + Exp) $\mathcal{T} = [0, 10]$; using kernel (K), projection (P), and numerical approximation (A) methods.

	(iv) Unif$[-1, 1]$			Exp(1)		
n	K	P	A	K	P	A
10	2.32	0.48	23.13	1.76	3.95	4.11
30	0.78	0.11	1.83	0.58	2.56	0.84
60	0.34	0.04	0.43	0.33	2.23	0.50
90	0.26	0.02	0.19	0.22	2.15	0.45

Table 2.2: Errors for three Unif$[-1, 1]$-distributed random variables estimated on $\mathcal{T} = [-5, 5]$ and three Exp(1) distributed random variables in $\mathcal{T} = [0, 10]$; using kernel (K), projection (P), and numerical approximation (A) methods with sample sizes $n = 10, 30, 60, 90$.

In Table 2.1 and Table 2.2 the performance of the kernel (K) and projection method (P) is compared to the numerical approximation (A) for different choices of sample size n and different densities. Here, for function $\widehat{f}$ evaluated at points $t_1, \ldots, t_N$ with interval steps Δ_t approximating function f we consider the error

$$100 * \sum_{i=1}^{N} \Delta_t \left(\widehat{f}(t_i) - f(t_i)\right)^2,$$

where we scale by 100 for better readability. For all examples, we chose $N = 100$ and $\{t_1, \ldots, t_N\}$ as the set of equidistant points on $\mathcal{T}$ depending on the example. For examples (i), (ii) and (iv) we set $\mathcal{T} = [-5, 5]$ and for (iii) we set $\mathcal{T} = [-2.5, 2.5]$. When considering estimation, we additionally take the mean over 50 Monte Carlo simulations. For the numerical approximation (A), the Fast Fourier Transform was used (in R implemented with the function `fft`) with n evaluation points.

As expected, increasing the sample size n reduces the error. However, there is not one method outperforming the others over all examples. The numerical approximation has smaller errors, e.g., for two exponentially distributed random variables. On the other hand, for the case of three exponentially distributed random variables, the kernel method performs best. Conversely, for two uniform distributed random variables the projection method appears to be most fitting. In addition, the relationships might change with increasing n. For instance, for three uniform distributed random variables, the kernel methods seems better than the numerical approximation for small n. But for large values the roles switch.

	Unif + Gaussian		Unif + Claw	
n	K	P	K	P
50	4.38	0.26	4.86	0.25
100	0.49	0.13	0.77	0.18
200	0.17	0.11	0.19	0.13

Table 2.3: Errors for estimating the distribution of the sum of a Unif$[-1,1]$ distributed variable and either a standard Gaussian or a Claw-distributed random variable. For both examples, the function was estimated on $\mathcal{T} = [-5, 5]$ and for the error FFT was used as the ground truth.

Lastly, in Table 2.3 the results are given for two examples where the analytical derivation of the convolution is not straightforward. To evaluate the methods, we compare the estimates with FFT taken as ground truth. Again, the results indicate a better performance of the projection method.

This concludes the simulation study and, with this, the discussion of the estimation of the additive convolution of multiple densities. Below we give auxiliary results used in this chapter.

2.4 Auxiliary results

The following result is an extension of Lemma 1 in Chesneau et al. (2013) and is used in several proofs.

Lemma 2.4.1:
Let $(u, v) \in \mathbb{C}^p \times \mathbb{C}^p$ such that $|u_j|, |v_j| \leq 1$ for all $j \in [\![p]\!]$. Then, for any integer $p \geq 1$, we have

$$\left| \prod_{j\in[\![p]\!]} u_j - \prod_{j\in[\![p]\!]} v_j \right| \leq \prod_{j\in[\![p]\!]} |u_j - v_j| + \sum_{j\in[\![p]\!]} C_j^p |u_j - v_j||v_{j+1}|$$

setting $v_{p+1} := v_1$, where $C_1^1 = 0$, $C_1^2 = C_2^2 = 1$ and for $p \geq 2$ we define $C_j^{p+1} := 3C_j^p$ for $j \in [\![p-1]\!]$, $C_p^{p+1} := C_p^p + 2^{p-1}$ and $C_{p+1}^{p+1} := 1 + 2C_p^p$. In addition, for $p \geq 2$ and $j \in [\![p]\!]$ it holds that $C_j^p \leq 3^{p-2}$.

Remark 2.4.2:
Similarly one shows that

$$\left| \prod_{j\in[\![p]\!]} u_j - \prod_{j\in[\![p]\!]} v_j - \prod_{j\in[\![p]\!]} (u_j - v_j) \right| \leq \sum_{j\in[\![p]\!]} C_j^p |u_j - v_j||v_{j+1}|.$$

Proof of Lemma 2.4.1. We proof both assertions simultaneously by induction.
Base case. For $p = 1$ this is obviously satisfied with $C_1^1 = 0$. For $p = 2$ we see for $(u, v) \in \mathbb{C}^2 \times \mathbb{C}^2$

that it holds

$$\begin{aligned}u_1u_2 - v_1v_2 &= (u_1 - v_1)(u_2 - v_2) + v_2u_2 + v_2u_1 - 2v_1v_2 \\ &= (u_1 - v_1)(u_2 - v_2) + v_2(u_1 - v_1) + v_1(u_2 - v_2).\end{aligned}$$

Consequently, by the triangle inequality, the inequality holds for $C_1^2 = C_2^2 = 1$. Further $C_1^2 = C_1^2 \leq 3^0 = 3^{2-2}$.

Induction step $p \mapsto p + 1$. Assume that the inequality holds for $p \geq 2$. Let $(u, v) \in \mathbb{C}^{p+1} \times \mathbb{C}^{p+1}$ with $|u_j|, |v_j| \leq 1$ for all $j \in [\![p + 1]\!]$. First, note that

$$\begin{aligned}&\prod_{j\in[\![p+1]\!]} u_j - \prod_{j\in[\![p+1]\!]} v_j = u_{p+1}\left(\prod_{j\in[\![p]\!]} u_j - \prod_{j\in[\![p]\!]} v_j\right) + (u_{p+1} - v_{p+1}) \prod_{j\in[\![p]\!]} v_j \\ = \;& (u_{p+1} - v_{p+1})\left(\prod_{j\in[\![p]\!]} u_j - \prod_{j\in[\![p]\!]} v_j\right) + v_{p+1}\left(\prod_{j\in[\![p]\!]} u_j - \prod_{j\in[\![p]\!]} v_j\right) + (u_{p+1} - v_{p+1}) \prod_{j\in[\![p]\!]} v_j.\end{aligned}$$

Taking the absolute we obtain with the triangle inequality

$$\begin{aligned}&\left|\prod_{j\in[\![p+1]\!]} u_j - \prod_{j\in[\![p+1]\!]} v_j\right| \\ &\leq |u_{p+1} - v_{p+1}|\left|\prod_{j\in[\![p]\!]} u_j - \prod_{j\in[\![p]\!]} v_j\right| + |v_{p+1}|\left|\prod_{j\in[\![p]\!]} u_j - \prod_{j\in[\![p]\!]} v_j\right| + |u_{p+1} - v_{p+1}| \prod_{j\in[\![p]\!]} |v_j|\end{aligned}$$

It follows using the inequality for p on $(u_1, \dots, u_p, v_1, \dots, v_p)$

$$\begin{aligned}&\left|\prod_{j\in[\![p+1]\!]} u_j - \prod_{j\in[\![p+1]\!]} v_j\right| \\ &\leq |u_{p+1} - v_{p+1}|\left(\prod_{j\in[\![p]\!]} |u_j - v_j| + \sum_{j\in[\![p]\!]} C_j^p |u_j - v_j||v_{j+1}|\right) \\ &\quad + |v_{p+1}|\left(\prod_{j\in[\![p]\!]} |u_j - v_j| + \sum_{j\in[\![p]\!]} C_j^p |u_j - v_j||v_{j+1}|\right) + |u_{p+1} - v_{p+1}| \prod_{j\in[\![p]\!]} |v_j|.\end{aligned}$$

Further, using that $|v_j|, |u_j| \leq 1$ for all $j \in [\![p + 1]\!]$ and reordering the terms leads to

$$\begin{aligned}&\left|\prod_{j\in[\![p+1]\!]} u_j - \prod_{j\in[\![p+1]\!]} v_j\right| \\ &\leq \prod_{j\in[\![p+1]\!]} |u_j - v_j| + 2\sum_{j\in[\![p-1]\!]} C_j^p |u_j - v_j||v_{j+1}| + 2C_p^p |u_{p+1} - v_{p+1}||v_1| \\ &\quad + 2^{p-1}|v_{p+1}||u_p - v_p| + \sum_{j\in[\![p-1]\!]} C_j^p |u_j - v_j||v_{j+1}| + C_p^p |v_{p+1}||u_p - v_p| \\ &\quad + |u_{p+1} - v_{p+1}||v_1|\end{aligned}$$

Finally, we get

$$
\begin{aligned}
&\left|\prod_{j\in[\![p+1]\!]} u_j - \prod_{j\in[\![p+1]\!]} v_j\right| \\
&= \prod_{j\in[\![p+1]\!]} |u_j - v_j| + \sum_{j\in[\![p-1]\!]} 3C_j^p |u_j - v_j||v_{j+1}| + \left(C_p^p + 2^{p-1}\right)|u_p - v_p||v_{p+1}| \\
&\quad + \left(1 + 2C_p^p\right)|u_{p+1} - v_{p+1}||v_1| \\
&= \prod_{j\in[\![p+1]\!]} |u_j - v_j| + \sum_{j\in[\![p+1]\!]} C_j^{p+1} |u_j - v_j||v_{j+1}|,
\end{aligned}
$$

where $C_j^{p+1} = 3C_j^p$ for $j \in [\![p-1]\!]$, $C_p^{p+1} = C_p^p + 2^{p-1}$ and $C_{p+1}^{p+1} = 1 + 2C_p^p$. For the second assertion, assume that $C_j^p \leq 3^{p-2}$ for p. Then it follows that

$$
\begin{aligned}
&C_j^{p+1} = 3C_j^p \leq 3 \cdot 3^{p-2} = 3^{(p+1)-2}, \quad \text{for } j \in [\![(p+1)-1]\!], \\
&C_p^{p+1} = C_p^p + 2^{p-1} \leq 3^{p-2} + 2^{p-1} = 3^{(p+1)-2}\left(\frac{1}{3} + \left(\frac{2}{3}\right)^{p-1}\right) \leq 3^{(p+1)-2} \\
&C_{p+1}^{p+1} = 1 + 2C_p^p \leq 1 + 2 \cdot 3^{p-2} \leq 3^{(p+1)-2}\left(3^{-p+1} + \frac{2}{3}\right) \leq 3^{(p+1)-2}.
\end{aligned}
$$

This concludes the proof. □

Lemma 2.4.3:
For $h_1 \in \mathbb{L}_d^2$ and $h_2 \in \mathbb{L}_d^1$ it holds that

$$
\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |h_1(\boldsymbol{x})h_1(\boldsymbol{y})h_2(\boldsymbol{x}+\boldsymbol{y})| d\lambda^d(\boldsymbol{x}) d\lambda^d(\boldsymbol{y}) \leq \|h_1\|_{\mathbb{L}_d^2}^2 \|h_2\|_{\mathbb{L}_d^1}.
$$

Proof of Lemma 2.4.3. Splitting the integral with respect to the variable and applying Cauchy–Schwarz inequality we get

$$
\begin{aligned}
&\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |h_1(\boldsymbol{x})h_1(\boldsymbol{y})h_2(\boldsymbol{x}+\boldsymbol{y})| d\lambda^d(\boldsymbol{x}) d\lambda^d(\boldsymbol{y}) \\
&= \int_{\mathbb{R}^d} |h_1(\boldsymbol{y})| \int_{\mathbb{R}^d} |h_1(\boldsymbol{x})||h_2(\boldsymbol{x}+\boldsymbol{y})|^{1/2} |h_2(\boldsymbol{x}+\boldsymbol{y})|^{1/2} d\lambda^d(\boldsymbol{y}) d\lambda^d(\boldsymbol{x}). \\
&\leq \int_{\mathbb{R}^d} |h_1(\boldsymbol{y})| \left(\int_{\mathbb{R}^d} |h_2(\boldsymbol{x}+\boldsymbol{y})| d\lambda^d(\boldsymbol{x})\right)^{1/2} \left(\int_{\mathbb{R}^d} |h_2(\boldsymbol{x}+\boldsymbol{y})||h_1(\boldsymbol{x})|^2 d\lambda^d(\boldsymbol{x})\right)^{1/2} d\lambda^d(\boldsymbol{y}).
\end{aligned}
$$

By change of variables and applying again Cauchy–Schwarz inequality holds that

$$
\begin{aligned}
&\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |h_1(\boldsymbol{x})h_1(\boldsymbol{y})h_2(\boldsymbol{x}+\boldsymbol{y})| d\lambda^d(\boldsymbol{x}) d\lambda^d(\boldsymbol{y}) \\
&\leq \|h_2\|_{\mathbb{L}_d^1}^{1/2} \int_{\mathbb{R}^d} |h_1(\boldsymbol{y})| \left(\int_{\mathbb{R}^d} |h_2(\boldsymbol{x}+\boldsymbol{y})||h_1(\boldsymbol{x})|^2 d\lambda^d(\boldsymbol{x})\right)^{1/2} d\lambda^d(\boldsymbol{y}). \\
&\leq \|h_2\|_{\mathbb{L}_d^1}^{1/2} \|h_1\|_{\mathbb{L}_d^2} \left(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |h_2(\boldsymbol{x}+\boldsymbol{y})||h_1(\boldsymbol{x})|^2 d\lambda^d(\boldsymbol{x}) d\lambda^d(\boldsymbol{y})\right)^{1/2}
\end{aligned}
$$

Finally, by rewriting the last line we get

$$\begin{aligned}&\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|h_1(\boldsymbol{x})h_1(\boldsymbol{y})h_2(\boldsymbol{x}+\boldsymbol{y})|d\lambda^d(\boldsymbol{x})d\lambda^d(\boldsymbol{y})\\&\le \|h_2\|_{\mathbb{L}^1_d}^{1/2}\|h_1\|_{\mathbb{L}^2_d}\|h_1\|_{\mathbb{L}^2_d}\|h_2\|_{\mathbb{L}^1_d}^{1/2}=\|h_1\|_{\mathbb{L}^2_d}^2\|h_2\|_{\mathbb{L}^1_d}.\end{aligned}$$

This gives the result and concludes the proof. □

Lemma 2.4.4:
Let

$$A:=\bigcup_{\tilde p\in[\![p-1]\!]}\left\{((k_1,l_1),\dots,(k_p,l_p))\in\times_{j\in[\![p]\!]}([\![n_j]\!]\times[\![n_j]\!]):\sum_{j\in[\![p]\!]}\mathbb{1}(k_j=l_j)=\tilde p\right\}.$$

There exists a constant $C_p>0$ such that

$$|A|\le(p-1)p!\left(\prod_{j\in[\![p]\!]}n_j^2\right)\sum_{j\in[\![p]\!]}\frac{1}{n_j}.$$

Proof. Denote

$$A_{\tilde p}:=\left\{((k_1,l_1),\dots,(k_p,l_p))\in\times_{j\in[\![p]\!]}([\![n_j]\!]\times[\![n_j]\!]):\sum_{j\in[\![p]\!]}\mathbb{1}(k_j=l_j)=\tilde p\right\}.$$

We first determine $|A_{\tilde p}|$ for fixed $\tilde p\in[\![p]\!]$. Note that for any index couple the amount of options of them being the same is

$$|\{(k_j,l_j)\in[\![n_j]\!]\times[\![n_j]\!]:k_j=l_j\}|=n_j,$$

and for them to be different is

$$|\{(k_j,l_j)\in[\![n_j]\!]\times[\![n_j]\!]:k_j\neq l_j\}|=n_j(n_j-1).$$

Consequently, for a fixed $\tilde p\in[\![p-1]\!]$ for $k_{j_1}=l_{j_1},\dots,k_{j_{\tilde p}}=l_{j_{\tilde p}}$ and the rest to be unequal for fixed indices $j_1,\dots,j_p$, there exist

$$\begin{aligned}\left(\prod_{m\in[\![\tilde p]\!]}n_{j_m}\right)\left(\prod_{m=\tilde p+1}^{p}n_{j_m}(n_{j_m}-1)\right)&\le\left(\prod_{j\in[\![p]\!]}n_j\right)\left(\prod_{m=\tilde p+1}^{p}n_{j_m}\right)\\&=\left(\prod_{j\in[\![p]\!]}n_j^2\right)\left(\prod_{m=\tilde p+1}^{p}\frac{n_{j_m}}{n_{j_m}}\right)\left(\prod_{m\in[\![\tilde p]\!]}\frac{1}{n_{j_m}}\right)\\&\le\left(\prod_{j\in[\![p]\!]}n_j^2\right)\sum_{j\in[\![p]\!]}\frac{1}{n_j}\end{aligned}\tag{2.49}$$

many possibilities. Note that the bound does not depend anymore on the choice of indices. Further, there are $\binom{p}{\tilde{p}} \leq p!$ many possibilities to choose indices $j_1, \ldots, j_p$ in this way. It follows that

$$|A_{\tilde{p}}| \leq p! \left(\prod_{j \in [\![p]\!]} n_j^2 \right) \sum_{j \in [\![p]\!]} \frac{1}{n_j}.$$

Finally, we get

$$|A| = \sum_{\tilde{p} \in [\![p]\!]} |A_{\tilde{p}}| \leq (p-1)p! \left(\prod_{j \in [\![p]\!]} n_j^2 \right) \sum_{j \in [\![p]\!]} \frac{1}{n_j}$$

which completes the proof. □

The following result is presented in Tsybakov (2008).

Lemma 2.4.5 (Generalized Minkowski inequality)**:**
For any Borel function $g \colon \mathbb{R}^2 \to \mathbb{R}$ we have

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(u,x) d\lambda(u) \right)^2 d\lambda(x) \leq \left[\int_{\mathbb{R}} \left(\int_{\mathbb{R}} g^2(u,x) d\lambda(x) \right)^{1/2} d\lambda(u) \right]^2.$$

The multivariate version of the inequality results from repetitively applying the first inequality, a proof is omitted.

Lemma 2.4.6 (Generalized Minkowski inequality – Multivariate)**:**
For $d \in \mathbb{N}$ any and any Borel function $g \colon \mathbb{R}^{2d} \to \mathbb{R}$ we have

$$\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} g(\boldsymbol{u},\boldsymbol{x}) d\lambda^d(\boldsymbol{u}) \right)^2 d\lambda^d(\boldsymbol{x}) \leq \left[\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} g^2(\boldsymbol{u},\boldsymbol{x}) d\lambda^d(\boldsymbol{x}) \right)^{1/2} d\lambda^d(\boldsymbol{u}) \right]^2.$$

CHAPTER 3

Quadratic functional estimation under multiplicative measurement errors

In this chapter, we consider inference for observations contaminated by some error. More precisely, we look at a multiplicative measurement error model. That is, we have access to an independent and identically distributed (i.i.d.) sample $(Y_j)_{j\in[\![n]\!]}$ of size $n \in \mathbb{N}$ from

$$Y = XU, \tag{3.1}$$

where X and U are strictly positive independent random variables. We assume that both random variables, X and U, admit a density with respect to the Lebesgue measure on $\mathbb{R}_+$. We denote the unknown density of X by f. Throughout this chapter, we assume that U admits a known density denoted by φ. Then, it follows that Y also admits a Lebesgue density g given by the multiplicative convolution of f and φ, i.e.

$$g(y) = (f \circledast \varphi)(y) := \int_{\mathbb{R}_+} f(x)\varphi(y/x)x^{-1}d\lambda(x).$$

Consequently, inference is based on observations of Y generated by a transformation $g = T(f)$ of the density f given by a multiplicative convolution with the error density φ. Throughout this chapter, we will use the terminology convolution to refer to the multiplicative convolution. The goal of this chapter is to consider for the multiplicative measurement error model quadratic functional estimation. That is, we want to estimate a quadratic transformation $\theta(f)$ of the density f. More precisely, we define the following quadratic functional for functions $f\colon \mathbb{R}_+ \to \mathbb{R}_{\geq 0}$ on the positive real line satisfying $f \in \mathbb{L}^2_+(\mathrm{x}^{2c-1}) \cap \mathbb{L}^1_+(\mathrm{x}^{c-1})$ with fixed $c \in \mathbb{R}$ given an arbitrary measurable symmetric density function $\mathrm{w}\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ we define

$$q^2(f) := \|\mathcal{M}_c[f]\|^2_{\mathbb{L}^2(\mathrm{w}^2)} = \int_{\mathbb{R}} |\mathcal{M}_c[f](t)|^2\, \mathrm{w}^2(t)d\lambda(t), \tag{3.2}$$

where $\mathcal{M}_c[f]$ denotes the Mellin transform of f, see Section 1.2 for the definition and basic properties. We propose an estimator of the quadratic functional $\theta(f) = q^2(f)$ given in Equation (3.2) of the density f based on the multiplicative convolution theorem, see Property 1.2.2, which yields

$$\mathcal{M}_c[g] = \mathcal{M}_c[f] \cdot \mathcal{M}_c[\varphi].$$

Further, we derive upper bounds for the quadratic risk of this estimator. One motivation for estimating quadratic functionals comes from testing theory: A quadratic functional of the difference between unknown density and null hypothesis provides an intuitive test statistic, which is considered in the subsequent chapter. It should also be mentioned that the test problem yields a lower bound for the estimation problem.

Related literature For the additive error model, Butucea (2007) and Schluttenhofer and Johannes (2020b,a) discuss extensively the link between testing theory and quadratic functional estimation. Quadratic functional estimation has received great attention in the past. In the case of direct observations, Bickel and Ritov (1988) are the first to discover a typical phenomenon in quadratic functional estimation in the density context: the so-called elbow effect. It refers to a sudden change in the behavior of the rates, as soon as the smoothness parameter crosses a critical threshold. Birgé and Massart (1995) look at minimax lower bounds concerning smooth functionals of the density. Donoho and Nussbaum (1990) discover a similar behavior in the Gaussian sequence space model, the smoothness assumptions being in this context replaced by geometric assumptions. Laurent (1996) shows minimax rates, where also more details on the density framework and alternative estimators can be found. Adaptive nonparametric quadratic functional estimation is considered in a Gaussian sequence model by Laurent and Massart (2000) and in a density context by Laurent (2005). In the case of indirect observations, quadratic functional estimation in an inverse Gaussian sequence space model is treated by Butucea and Meziani (2011) for a known operator and in Kroll (2019) for partially unknown operators. For the additive measurement error model on the real line, quadratic functional estimation is considered, for example, by Chesneau (2011), Meister (2009) and Butucea (2007). Circular observations are studied in Schluttenhofer and Johannes (2020b,a).

In this work we extend quadratic functional estimation to the multiplicative measurement error model. Multiplicative censoring, which corresponds to the multiplicative measurement error model with U being uniformly distributed on $[0, 1]$, has been introduced and studied by Vardi (1989) and Vardi and Zhang (1992). van Es et al. (2000) explain and motivate the use of multiplicative censoring models in survival analysis. Concerning inference in this model, there has been extensive work on estimation of the unknown density f. To name a few: Andersen and Hansen (2001) consider series expansion methods treating the model as an inverse problem, Brunel et al. (2016) use a kernel estimator for density estimation in the multiplicative censoring model, Comte and Dion (2016) consider a projection density estimator with respect to the Laguerre basis, Belomestny et al. (2016) study a Beta-distributed error. The multiplicative measurement error model covers all of these cases of multiplicative censoring. Nonparametric density estimation in the multiplicative measurement error model has been considered by Brenner Miguel et al. (2021) using a spectral cut-off regularization. Brenner Miguel (2022) considers an estimation procedure using an anisotropic spectral cut-off. Brenner Miguel et al. (2023) look at the estimation of a linear functional of the unknown density and Belomestny and Goldenshluger (2020) examine point-wise density estimation in the multiplicative measurement error model.

Contribution We now turn to the nonparametric estimation of the value of a weighted quadratic functional evaluated at the Mellin transform of the unknown density in a multiplicative measurement error model. Thereby, our work extends the results of the estimation of the value of a quadratic functional evaluated at the density to the multiplicative measurement error model and, further, its generalized formulation covers not only the estimation the (possibly weighted) $\mathbb{L}^2$-norm of the density but also the $\mathbb{L}^2$-norm of its derivatives or the $\mathbb{L}^2$-norm of its survival function. More precisely, by applying a Plancherel type identity those examples can be written as the value of a weighted quadratic functional evaluated at the Mellin transform of the unknown density. Following the estimation strategy in Brenner Miguel et al. (2021) we define a spectral cut-off estimator of the unknown density f which we plug-in the quadratic functional. The accuracy of the estimator is measured by its mean squared error and it depends crucially on the cut-off parameter. Our aim is to establish a fully data-driven estimation procedure inspired by Goldenshluger and Lepski (2011) and to derive upper bounds for its mean squared error as well as convergence rates.

Outline of this chapter This chapter is organized as follows. We start in Section 3.1 by presenting the estimation strategy. In Section 3.2 we first decompose the estimation error of the plug-in estimator appropriately and then derive an upper bound of its mean squared error. Introducing Mellin-Sobolev spaces we derive convergence rates resulting from the upper bound in Section 3.3. Furthermore, we introduce a data-driven procedure based on the Goldenshluger-Lepski method, show an upper bound for its mean squared error and calculate convergence rates in Section 3.4. Finally, we illustrate our results by a short simulation study in Section 3.5.

3.1 Estimation strategy

We aim to estimate the value of the quadratic functional $q^2 := q^2(f)$ defined in Equation (3.2) of the density f in the multiplicative measurement error model. All needed properties for the estimation problem and the below proposed estimator to be well-defined are summarized in the following assumption.

Assumption 3.1.1:
Consider the multiplicative measurement error model Equation (3.1), an arbitrary measurable symmetric density function $\mathrm{w}\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ and $c \in \mathbb{R}$. In addition, let

(i) $f, \varphi \in \mathbb{L}^2_+(\mathrm{x}^{2c-1})$,

(ii) $\mathcal{M}_c[f] \in \mathbb{L}^2(\mathrm{w}^2)$,

(iii) $\mathcal{M}_c[\varphi](t) \neq 0$ for all $t \in \mathbb{R}$ and $\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi] \in \mathbb{L}^\infty(\mathrm{w})$ for all $k \in \mathbb{R}_+$.

(iv) $f, \varphi \in \mathbb{L}^1_+(\mathrm{x}^{2c-2})$ and $\varphi \in \mathbb{L}^\infty_+(\mathrm{x}^{2c-1})$.

Let us discuss the given assumptions in detail. With Assumption 3.1.1 (i) the existence of the Mellin transform of f and φ is ensured. Note that since $g = f \circledast \varphi$ it follows $g \in \mathbb{L}^2_+(\mathrm{x}^{2c-1})$, see Property 1.2.3, and its Mellin transform exists as well. Under Assumption 3.1.1 (ii) the parameter q^2 of interest is well-defined, see Equation (3.2). With additionally Assumption 3.1.1 (iii) the estimator $\hat{q}^2_k$ proposed below in Equation (3.4) is well-defined, too.

Further, it follows that $\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi] \in \mathbb{L}^2(\mathrm{w}^2) \cap \mathbb{L}^4(\mathrm{w}^4)$ since by Cauchy–Schwarz inequality it holds that

$$\begin{aligned}\left(\int_{-k}^{k} \frac{1}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t)\right)^{1/2} &\leq (2k)^{1/4}\left(\int_{-k}^{k} \frac{1}{|\mathcal{M}_c[\varphi](t)|^4}\,\mathrm{w}^4(t)d\lambda(t)\right)^{1/4}\\ &\leq (2k)^{1/2}\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^\infty(\mathrm{w})} < \infty.\end{aligned}$$

Moreover, it holds for all $t \in \mathbb{R}$

$$\mathbb{E}_\varphi[U^{c-1}] = \mathbb{E}_\varphi[|U^{c-1+2\pi it}|] \geq |\mathcal{M}_c[\varphi](t)|.$$

It follows $\|\mathcal{M}_c[\varphi]\mathbb{1}_{[-k,k]}\|^2_{\mathbb{L}^2} \leq 2k(\mathbb{E}_\varphi[U^{c-1}])^2$ and, hence,

$$\|\mathbb{1}_{[-k,k]}\|^2_{\mathbb{L}^2(\mathrm{w}^2)} \leq \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^\infty(\mathrm{w})}(2k)(\mathbb{E}_\varphi[U^{c-1}])^2.$$

Consequently, Assumption 3.1.1 (iii) also ensures that $\mathbb{1}_{[-k,k]} \in \mathbb{L}^2(\mathrm{w}^2)$. If in (iii) the condition $\mathcal{M}_c[\varphi](t) \neq 0$ is not satisfied, one could consider a different value of the development point $c \in \mathbb{R}$. Changing the development point is similar in spirit to the approach presented in Belomestny and Goldenshluger (2021), where the estimation procedure within the additive deconvolution problem on the real line is analyzed on a strip as a subset of the complex plane $\mathbb{C}$. Under Assumption 3.1.1 (iv), it follows that $g \in \mathbb{L}^1_+(\mathrm{x}^{2c-2}) \cap \mathbb{L}^\infty_+(\mathrm{x}^{2c-1})$ and $\varphi \in \mathbb{L}^1_+(\mathrm{x}^{c-1})$, hence, with (i) it is possible to apply the multiplicative convolution theorem Property 1.2.2. Note that square integrability is a common assumption in additive deconvolution, which might be here seen as the particular case $c = \frac{1}{2}$ and $\mathrm{w} = 1$, see Example 1.4.1. However, allowing for different values $c \in \mathbb{R}$ makes the dependence on the assumptions visible.

For the construction of the estimator, we intend to plug-in an estimator of the Mellin transform of the unknown density f into the norm expression given in Equation (3.2). We follow the approach of Brenner Miguel et al. (2021) for density estimation in the multiplicative measurement error model. Using the multiplicative convolution theorem Property 1.2.2, we have that $\mathcal{M}_c[f] = \mathcal{M}_c[g]/\mathcal{M}_c[\varphi]$. Recall that the error density is assumed to be known. Consequently, to get an estimator of q^2 one would be tempted to plug in Equation (3.2) an estimator for the unknown Mellin transform $\mathcal{M}_c[g]$. A natural choice is the empirical Mellin transform

$$\widehat{\mathcal{M}}_c(t) := \frac{1}{n}\sum_{j\in[\![n]\!]} Y_j^{c-1+2\pi it}, \quad t \in \mathbb{R}.$$

Observe that $\widehat{\mathcal{M}}_c$ is a point-wise unbiased estimator of $\mathcal{M}_c[g]$, i.e., by the definition of the Mellin transform, see Equation (1.4), it holds for any $t \in \mathbb{R}$ that

$$\mathbb{E}^n_g[\widehat{\mathcal{M}}_c(t)] = \mathbb{E}_g[Y_1^{c-1+2\pi it}] = \mathcal{M}_c[g](t).$$

Consequently, we have for the point-wise scaled variance of $\widehat{\mathcal{M}}_c$ that for each $t \in \mathbb{R}$ using the independence of observations $(Y_j)_{j\in[\![n]\!]}$ that

$$\begin{aligned} n\mathbb{E}^n_g[|\widehat{\mathcal{M}}_c(t) - \mathcal{M}_c[g](t)|^2] &= \frac{1}{n}\sum_{l,j\in[\![n]\!]} \mathbb{E}[(Y_j^{c-1+2\pi it} - \mathcal{M}_c[g](t))(Y_l^{c-1-2\pi it} - \mathcal{M}_c[g](-t))] \\ &= \mathbb{E}_g[|Y_1^{c-1+2\pi it} - \mathcal{M}_c[g](t)|^2] = \mathrm{var}(Y_1^{c-1+2\pi it}) \end{aligned}$$

and, thus, it follows

$$\begin{aligned} n\mathbb{E}^n_g[|\widehat{\mathcal{M}}_c(t) - \mathcal{M}_c[g](t)|^2] &= \mathrm{var}(Y_1^{c-1+2\pi it}) \leq \mathbb{E}[|Y_1^{c-1+2\pi it}|^2] \\ &= \mathbb{E}_g[Y_1^{2c-2}] = \int_{\mathbb{R}_+} g(y)y^{2c-2}d\lambda_+(y) = \|g\|_{\mathbb{L}^1_+(\mathrm{x}^{2c-2})}. \end{aligned} \tag{3.3}$$

However, $\widehat{\mathcal{M}}_c/\mathcal{M}_c[\varphi]$ is in general not square integrable. A frequently used technique in inverse problems is the so called spectral cut-off method, see for example Engl et al. (1996). Applying this method to the estimator and plugging it into Equation (3.2) results for some $k \in \mathbb{R}_+$ in the term

$$\left\| \mathbb{1}_{[-k,k]} \frac{\widehat{\mathcal{M}}_c}{\mathcal{M}_c[\varphi]} \right\|^2_{\mathbb{L}^2(\mathrm{w}^2)}.$$

Correcting for the bias finally yields the estimator

$$\hat{q}_k^2 := \frac{1}{n(n-1)} \sum_{\substack{j \neq l \\ j,l \in [\![n]\!]}} \int_{-k}^{k} \frac{Y_j^{c-1+2\pi i t} Y_l^{c-1-2\pi i t}}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t). \tag{3.4}$$

Using the independence of $(Y_j)_{j \in [\![n]\!]}$ and Property 1.2.3 (ii), namely $\overline{\mathcal{M}_c[g]}(t) = \mathcal{M}_c[g](-t)$ for all $t \in \mathbb{R}$, and the convolution theorem Property 1.2.2, we have that

$$\begin{aligned} q_k^2 := \mathbb{E}_g^n[\hat{q}_k^2] &= \int_{-k}^{k} \frac{\mathbb{E}_g[Y_j^{c-1+2\pi i t}] \mathbb{E}_g[Y_l^{c-1-2\pi i t}]}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t) \\ &= \int_{-k}^{k} \frac{|\mathcal{M}_c[g](t)|^2}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t) = \int_{-k}^{k} |\mathcal{M}_c[f](t)|^2 \mathrm{w}^2(t) d\lambda(t). \end{aligned} \tag{3.5}$$

3.2 Upper bound for the quadratic risk

In this section, we derive an upper bound for the mean squared error of the spectral cut-off estimator $\hat{q}_k^2$ defined in Equation (3.4). For this, we first introduce some notation. Define for $k \in \mathbb{R}_+$

$$\mathrm{v}_k := \int_{-k}^{k} \frac{1}{|\mathcal{M}_c[\varphi](t)|^4} \mathrm{w}^4(t) d\lambda(t), \tag{3.6}$$

$$\Lambda_{f|\varphi}(k) := \int_{-k}^{k} \frac{|\mathcal{M}_c[f](t)|^2}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^4(t) d\lambda(t). \tag{3.7}$$

For $x, y \in \mathbb{R}$ denote $x \vee y := \max\{x, y\}$ and $x \wedge y := \min\{x, y\}$. Further define

$$\mathrm{v}_{f|\varphi} := (\|g\|_{\mathbb{L}^1_+(\mathrm{x}^{2(c-1)})} \vee 1) \quad \text{and} \quad \mathrm{C}_\varphi := (\|\varphi\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})} / \|\varphi\|_{\mathbb{L}^1_+(\mathrm{x}^{2c-2})} \vee 1), \tag{3.8}$$

which are finite under Assumption 3.1.1. Moreover, by Lemma 1.2.5 (iii), it holds

$$\|g\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})} \leq \frac{\|\varphi\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})}}{\|\varphi\|_{\mathbb{L}^1_+(\mathrm{x}^{2c-2})}} \|g\|_{\mathbb{L}^1_+(\mathrm{x}^{2(c-1)})} \leq \mathrm{v}_{f|\varphi} \mathrm{C}_\varphi, \tag{3.9}$$

and, thus,

$$\|g\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})} \|g\|_{\mathbb{L}^1_+(\mathrm{x}^{2c-2})} \leq \mathrm{C}_\varphi \mathrm{v}_{f|\varphi}^2. \tag{3.10}$$

Both inequalities will be used in the proofs of this chapter. We get the following result for the quadratic risk.

Proposition 3.2.1:
Under Assumption 3.1.1 the estimator $\hat{q}_k^2$ defined in Equation (3.4) for q^2 satisfies for $k \in \mathbb{R}_+$ and $n \geq 2$

$$\mathbb{E}_g^n[|\hat{q}_k^2 - q^2|^2] \leq (q_k^2 - q^2)^2 + 4\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}^2 \left(\frac{\mathbf{v}_k}{n^2} + \frac{\Lambda_{f|\varphi}(k)}{n} \right).$$

Proof of Proposition 3.2.1. The main step of the proof is to decompose the estimation error $\hat{q}_k^2 - q^2$ into a U-statistic and a linear statistics. More precisely, introduce the U-statistic U_k given for $k \in \mathbb{R}_+$ by

$$U_k := \frac{1}{n(n-1)} \sum_{\substack{j \neq l \\ j,l \in [\![n]\!]}} \int_{-k}^{k} \frac{(Y_j^{c-1+2\pi it} - \mathcal{M}_c[g](t))(Y_l^{c-1-2\pi it} - \mathcal{M}_c[g](-t))}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t) \quad (3.11)$$

and the linear statistic W_k given by

$$W_k := \frac{1}{n} \sum_{j \in [\![n]\!]} \int_{-k}^{k} (Y_j^{c-1+2\pi it} - \mathcal{M}_c[g](t)) \frac{\mathcal{M}_c[f](-t)}{\mathcal{M}_c[\varphi](t)} \mathrm{w}^2(t) d\lambda(t). \quad (3.12)$$

With this, we obtain for q^2 defined in Equation (3.2), $\hat{q}_k^2$ in Equation (3.4) and q_k^2 in Equation (3.5) the decomposition

$$\hat{q}_k^2 - q^2 = U_k + 2W_k - (q_k^2 - q^2). \quad (3.13)$$

The statistics U_k and W_k are centered since the observations are independent and for all $j \in [\![n]\!]$ it holds $\mathbb{E}_g[Y_j^{c-1+2\pi it} - \mathcal{M}_c[g](t)] = 0$. Further the statistics are uncorrelated. Indeed, it holds with the same argument that

$$\begin{aligned}
\mathbb{E}_g^n[U_k W_k] &= \frac{1}{n^2(n-1)} \sum_{\substack{j \neq l \\ j,l \in [\![n]\!]}} \sum_{m \in [\![n]\!]} \mathbb{E}_g^3 \Big[\Big(\int_{-k}^{k} (Y_m^{c-1+2\pi it} - \mathcal{M}_c[g](t)) \frac{\mathcal{M}_c[f](-t)}{\mathcal{M}_c[\varphi](t)} \mathrm{w}^2(t) d\lambda(t) \Big) \\
&\quad \cdot \Big(\int_{-k}^{k} \frac{(Y_j^{c-1+2\pi it} - \mathcal{M}_c[g](t))(Y_l^{c-1-2\pi it} - \mathcal{M}_c[g](-t))}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t) \Big) \Big] \\
&= \frac{2}{n-1} \sum_{\substack{l \neq 1 \\ j \in [\![n]\!]}} \mathbb{E}_g^3 \Big[\Big(\int_{-k}^{k} (Y_1^{c-1+2\pi it} - \mathcal{M}_c[g](t)) \frac{\mathcal{M}_c[f](-t)}{\mathcal{M}_c[\varphi](t)} \mathrm{w}^2(t) d\lambda(t) \Big) \\
&\quad \cdot \Big(\int_{-k}^{k} \frac{(Y_1^{c-1+2\pi it} - \mathcal{M}_c[g](t))(Y_l^{c-1-2\pi it} - \mathcal{M}_c[g](-t))}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t) \Big) \Big] = 0.
\end{aligned}$$

Thus, we obtain the decomposition of the quadratic risk

$$\mathbb{E}_g^n \left[\left| \hat{q}_k^2 - q^2 \right|^2 \right] = \mathbb{E}_g^n[|U_k|^2] + 4\mathbb{E}_g^n[|W_k|^2] + (q_k^2 - q^2)^2. \quad (3.14)$$

We first consider U_k defined in Equation (3.11) and show that it is a canonical U-statistic. Define for $x, y \in \mathbb{R}_+$ the function

$$h_k(x, y) := \int_{-k}^{k} \frac{(x^{c-1+2\pi it} - \mathcal{M}_c[g](t))(y^{c-1-2\pi it} - \mathcal{M}_c[g](-t))}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t). \quad (3.15)$$

The function h_k is symmetric. More precisely, by substitution of variables and since w is assumed to be symmetric it holds that

$$h_k(x,y) = \int_{-k}^{k} \frac{(x^{c-1-2\pi it} - \mathcal{M}_c[g](-t))(y^{c-1+2\pi it} - \mathcal{M}_c[g](t))}{|\mathcal{M}_c[\varphi](-t)|^2} \mathrm{w}^2(-t)d\lambda(t) = h_k(y,x).$$

Similarly, it holds that $\overline{h_k(x,y)} = h_k(y,x) = h_k(x,y)$, i.e., h_k is real-valued. Since h_k is symmetric and for all $y \in \mathbb{R}_+$ it holds that

$$\begin{aligned}\mathbb{E}_g[h_k(y,Y_1)] &= \mathbb{E}_g[h_k(Y_1,y)] \\ &= \int_{-k}^{k} \frac{\mathbb{E}_g[Y_1^{c-1+2\pi it} - \mathcal{M}_c[g](t)](y^{c-1-2\pi it} - \mathcal{M}_c[g](-t))}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t)d\lambda(t) = 0.\end{aligned}$$

Then, $U_k = \frac{1}{n(n-1)} \sum_{\substack{j \neq l \\ j,l \in [\![n]\!]}} h_k(Y_j, Y_l)$ is a canonical U-statistic. Using that U_k is canonical we get by Lemma 3.6.1 given in Section 3.6

$$\mathrm{Var}(U_k) \leq \frac{1}{n(n-1)} \mathbb{E}_g^2[h_k^2(Y_1,Y_2)].$$

For every $y \in \mathbb{R}_+$ using the property of the Mellin transform given in Lemma 1.2.5 (i) with $h(t) := \frac{(y^{c-1-2\pi it} - \mathcal{M}_c[g](-t))}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t)$ we have

$$\begin{aligned}\mathbb{E}_g[|h_k(Y_1,y)|^2] &\leq \mathbb{E}_g\left[\left|\int_{-k}^{k} Y_1^{c-1+2\pi it} \frac{(y^{c-1-2\pi it} - \mathcal{M}_c[g](-t))}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t)d\lambda(t)\right|^2\right] \\ &\leq \|g\|_{\mathbb{L}_+^\infty(\mathrm{x}^{2c-1})} \int_{-k}^{k} |y^{c-1-2\pi it} - \mathcal{M}_c[g](-t)|^2 \frac{\mathrm{w}^4(t)}{|\mathcal{M}_c[\varphi](t)|^4} d\lambda(t).\end{aligned}$$

Hence, applying this bound and the variance bound of the empirical Mellin transform derived in Equation (3.3) we have that

$$\begin{aligned}\mathbb{E}_g^2[h_k^2(Y_1,Y_2)] &\leq \|g\|_{\mathbb{L}_+^\infty(\mathrm{x}^{2c-1})} \int_{-k}^{k} \mathbb{E}_g\left[|Y_1^{c-1-2\pi it} - \mathcal{M}_c[g](-t)|^2\right] \frac{\mathrm{w}^4(t)}{|\mathcal{M}_c[\varphi](t)|^4} d\lambda(t) \\ &\leq \|g\|_{\mathbb{L}_+^\infty(\mathrm{x}^{2c-1})} \|g\|_{\mathbb{L}_+^1(\mathrm{x}^{2c-2})} \int_{-k}^{k} \frac{\mathrm{w}^4(t)}{|\mathcal{M}_c[\varphi](t)|^4} d\lambda(t).\end{aligned}$$

With Equation (3.10), for all $n \geq 2$ it follows

$$\mathbb{E}_g^n[|U_k|^2] \leq \frac{1}{n(n-1)} \mathrm{C}_\varphi \mathrm{v}_{f|\varphi}^2 \mathbf{v}_k \leq \frac{2}{n^2} \mathrm{C}_\varphi \mathrm{v}_{f|\varphi}^2 \mathbf{v}_k. \tag{3.16}$$

Let us turn to W_k. Recall that W_k is centered. Then, using in addition the Mellin property given in Lemma 1.2.5 (i) we get that

$$\begin{aligned}n\mathbb{E}_g^n[|W_k|^2] = n\,\mathrm{Var}(W_k) &= \mathrm{Var}\left(\int_{-k}^{k} Y_1^{c-1+2\pi it} \frac{\mathcal{M}_c[f](-t)}{\mathcal{M}_c[\varphi](t)} \mathrm{w}^2(t)d\lambda(t)\right) \\ &\leq \mathbb{E}_g\left[\left|\int_{-k}^{k} Y_1^{c-1+2\pi it} \frac{\mathcal{M}_c[f](-t)}{\mathcal{M}_c[\varphi](t)} \mathrm{w}^2(t)d\lambda(t)\right|^2\right] \\ &\leq \|g\|_{\mathbb{L}_+^\infty(\mathrm{x}^{2c-1})} \int_{-k}^{k} \frac{|\mathcal{M}_c[f](t)|^2}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^4(t)d\lambda(t),\end{aligned}$$

which together with Equation (3.9) implies

$$\mathbb{E}_g^n[|W_k|^2] \leq \mathrm{C}_\varphi \mathrm{v}_{f|\varphi} \frac{\Lambda_{f|\varphi}(k)}{n}. \tag{3.17}$$

Plugging the bound from Equation (3.16) and Equation (3.17) into Equation (3.14) we get

$$\mathbb{E}_g^n\left[\left|\hat{q}_k^2 - q^2\right|^2\right] \leq (q_k^2 - q^2)^2 + \frac{2}{n^2}\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}^2 \mathbf{v}_k + 4\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\frac{\Lambda_{f|\varphi}(k)}{n}.$$

Since $\mathrm{v}_{f|\varphi} \geq 1$, this completes the proof. □

Remark 3.2.2:
Let us briefly discuss Proposition 3.2.1. First, note that by definition $\mathbf{v}_k = \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^4(\mathrm{w}^4)}$ and $\Lambda_{f|\varphi} = \|\mathbb{1}_{[-k,k]}\mathcal{M}_c[f]/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^2(\mathrm{w}^4)}$ are monotonically increasing in k while the bias term

$$(q_k^2 - q^2)^2 = \left(\int_{\mathbb{R}\setminus[-k,k]} |\mathcal{M}_c[f](t)|^2\, \mathrm{w}^2(t) d\lambda(t)\right)^2$$

is monotonically decreasing in k. Consequently, for an optimal choice of k, one has to balance these terms, known as the classical bias-variance trade-off. In the case when both $\Lambda_{f|\varphi}$ and $\mathbf{v}_k$ are bounded, we retain a parametric rate. Note also that an optimal choice of k depends on n and the unknown density f. This motivates a data-driven choice, which we discuss below in Section 3.4 in more detail.

Next, we discuss uniform boundedness of the upper bound of Proposition 3.2.1 over a nonparametric class of functions to derive convergence rates.

3.3 Rates of convergence

From the upper bound result Proposition 3.2.1, we now derive rates for specific function classes for the three examples for parameter q^2 given in Section 1.4.3 depending on some function s satisfying the following assumptions.

Assumption 3.3.1: (i) Let $\mathrm{s}\colon \mathbb{R} \to \mathbb{R}$ be a symmetric, monotonically increasing weight function such that $\mathrm{s}(t) \to \infty$ as $|t| \to \infty$.

(ii) Assume that the symmetric function w/s is monotonically non-increasing such that $\mathrm{w}(t)/\mathrm{s}(t) = o(1)$ as $|t| \to \infty$.

For s satisfying Assumption 3.3.1 and $L > 0$ we set

$$\mathscr{F}(\mathrm{s}, L) := \{f \in \mathbb{L}^2_+(\mathrm{x}^{2c-1}) : \|\mathrm{s}\,\mathcal{M}_c[f]\|^2_{\mathbb{L}^2} \leq L^2, \|f\|_{\mathbb{L}^1_+(\mathrm{x}^{2c-2})} \leq L^2\}. \tag{3.18}$$

Roughly speaking, s describes how fast the Mellin transform $\mathcal{M}_c[f](t)$ vanishes for increasing values $|t|$. Note that we consider regularity classes independent of density function w, which specifies the quadratic functional of interest, e.g., the $\mathbb{L}^2_+$-norm of the density itself, its survival function or some derivative of the density, see Examples 1.4.1 to 1.4.3.

With Assumption 3.3.1 (i) the class $\mathscr{F}(\mathrm{s}, L)$ covers the usual assumptions on the regularity of the unknown density f, i.e., ordinary and super smooth densities. More precisely, for two real-valued functions $h_1, h_2 \colon \mathbb{R} \to \mathbb{R}$ we write $h_1(t) \sim h_2(t)$ if there exist constants $C, \tilde{C}$ such that for all $t \in \mathbb{R}$ we have $h_1(t) \le C h_2(t)$ and $h_2(t) \le \tilde{C} h_1(t)$. We then call h_2 the order of h_1. Then, in the *ordinary smooth case* the weight function s is of polynomial order, i.e., $\mathrm{s}(t) \sim (1+t^2)^{s/2}$ for $t \in \mathbb{R}$ and some $s \in \mathbb{R}_+$. In the *super smooth* case s is assumed to be exponentially increasing, i.e., $\mathrm{s}(t) \sim \exp(|t|^s)$ for $t \in \mathbb{R}$ some $s \in \mathbb{R}_+$. In the ordinary smooth case, $\mathscr{F}(\mathrm{s}, L)$ corresponds to the Mellin-Sobolev space, see Definition 2.3.9, Brenner Miguel (2023). Concerning Assumption 3.3.1 (ii), boundedness of $\mathrm{w}\,/\,\mathrm{s}$ ensures that the quadratic functional is finite on the nonparametric class of functions $\mathscr{F}(\mathrm{s}, L)$. Convergence against zero gives the rate for the bias term, as can be seen in the next result Corollary 3.3.2.

If $\mathrm{w}(t) \sim t^a$ for $a \in \mathbb{R}$, this includes the examples Examples 1.4.1 to 1.4.3 with the following choices of a.

(i) In the case $a = 0$ we have $\mathrm{w}(t) = 1$ for all $t \in \mathbb{R}$. Recall that in this case q^2 equals a (weighted) $\mathbb{L}^2_+$-norm of the density f itself.

(ii) The case $a = -1$ covers quadratic functional estimation of the survival function S_f of X, more precisely, we have $\mathrm{w}^2(t) = \frac{1}{(c-1)^2 + 4\pi t^2}$ for all $t \in \mathbb{R}$.

(iii) The case $a = \beta \in \mathbb{N}$ covers quadratic functionals of derivatives $D^\beta[f]$.

In addition, in case of ordinary smoothness of the unknown density f Assumption 3.3.1 (ii) imposes the condition $s > a$ on the parameters. Otherwise all the examples satisfy Assumption 3.3.1 for t large enough for both the case of ordinary smoothness and super smoothness.

For the nonparametric class $\mathscr{F}(\mathrm{s}, L)$ of functions we first derive a uniform bound for the risk to derive convergence rates. For this define for $k \in \mathbb{R}_+$ the term

$$\mathbf{m}_k := \left\| \frac{\mathrm{w}\,\mathbb{1}_{[-k,k]}}{\mathcal{M}_c[\varphi]} \right\|^4_{\mathbb{L}^\infty}. \tag{3.19}$$

Note that $\mathbf{v}_k = \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^4(\mathrm{w}^4)} \le (2k)\mathbf{m}_k$. Consequently, depending on the behavior of $\mathcal{M}_c[\varphi]$ either term can dominate the other. Further, let

$$R_n(k) := \frac{\mathrm{w}^4(k)}{\mathrm{s}^4(k)} \vee \frac{\mathbf{m}_k \vee \mathbf{v}_k}{n^2} \text{ and } R_n^{\mathrm{elbow}} := \left\| \frac{\mathrm{w}^4}{\mathrm{s}^4} \wedge \frac{\mathrm{w}^4}{\mathrm{s}^2\, n |\mathcal{M}_c[\varphi]|^2} \right\|_{\mathbb{L}^\infty}. \tag{3.20}$$

Corollary 3.3.2:
Under Assumption 3.1.1 and Assumption 3.3.1, if $f \in \mathscr{F}(\mathrm{s}, L)$, see Equation (3.18), then the estimator $\hat{q}^2_k$ defined in Equation (3.4) satisfies for $n \in \mathbb{N}$, $n \ge 2$ and $k \in \mathbb{R}_+$ that

$$\sup_{f \in \mathscr{F}(\mathrm{s}, L)} \mathbb{E}^n_g[|\hat{q}^2_k - q^2|^2] \le C(R_n(k) \vee R_n^{\mathrm{elbow}}), \tag{3.21}$$

for some constant $C > 0$ depending on φ and L.

Let us briefly comment the last result before we state the proof.

Remark 3.3.3:
Note that the term $R_n(k)$ in the upper bound in Corollary 3.3.2 depends on the parameter k. On the other hand, term R_n^{elbow} does not depend on k and, as becomes more clear in Table 3.1, this causes the so-called elbow effect. This also shows that, if $\mathbf{m}_k$ and $\mathbf{v}_k$ are bounded in k, a parametric rate is retained but otherwise not and that the choice of k needs to depend on n.

Proof of Corollary 3.3.2. From Proposition 3.2.1 for $f \in \mathscr{F}(\mathrm{s}, L)$ and $\mathrm{w}\,/\,\mathrm{s}$ increasing follows

$$\begin{aligned}(q_k^2 - q^2)^2 &= \|\mathbb{1}_{\mathbb{R}\setminus[-k,k]}\mathcal{M}_c[f]\|^4_{\mathbb{L}^2(\mathrm{w}^2)} = \left\|\frac{\mathrm{s}}{\mathrm{s}}\mathbb{1}_{\mathbb{R}\setminus[-k,k]}\mathcal{M}_c[f]\right\|^4_{\mathbb{L}^2(\mathrm{w}^2)} \\ &\leq \|\,\mathrm{s}\,\mathcal{M}_c[f]\|^2_{\mathbb{L}^2}\left\|\mathbb{1}_{\mathbb{R}\setminus[-k,k]}\frac{\mathrm{w}^4}{\mathrm{s}^4}\right\|_{\mathbb{L}^\infty} \leq L^4\left\|\mathbb{1}_{\mathbb{R}\setminus[-k,k]}\frac{\mathrm{w}^4}{\mathrm{s}^4}\right\|_{\mathbb{L}^\infty} \leq L^4\frac{\mathrm{w}^4(k)}{\mathrm{s}^4(k)}.\end{aligned}$$

To control $\Lambda_{f|\varphi}$ first note that if $n|\mathcal{M}_c[\varphi](t)|^2 \geq \mathrm{s}^2(t)$

$$\frac{1}{n}\frac{|\mathcal{M}_c[f](t)|^2}{|\mathcal{M}_c[\varphi](t)|^2} \leq |\mathcal{M}_c[f](t)|^2\,\mathrm{s}^2(t)\left(\frac{1}{\mathrm{s}^4(t)} \wedge \frac{1}{\mathrm{s}^2(t)n|\mathcal{M}_c[\varphi](t)|^2}\right)$$

and if $n|\mathcal{M}_c[\varphi](t)|^2 < \mathrm{s}^2(t)$

$$\frac{1}{n}\frac{|\mathcal{M}_c[f](t)|^2}{|\mathcal{M}_c[\varphi](t)|^2} \leq |\mathcal{M}_c[f](t)|^2\,\mathrm{s}^2(t)\left(\frac{1}{n^2|\mathcal{M}_c[\varphi](t)|^4} \wedge \frac{1}{s^2(t)n|\mathcal{M}_c[\varphi](t)|^2}\right).$$

Hence, we obtain for the entire integral

$$\begin{aligned}\Lambda_{f|\varphi}(k) &= \frac{1}{n}\int_{-k}^{k}\frac{|\mathcal{M}_c[f](t)|^2}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^4(t)d\lambda(t) \\ &\leq \int_{-k}^{k}|\mathcal{M}_c[f](t)|^2\,\mathrm{s}^2(t)\left(\frac{\mathrm{w}^4(t)}{\mathrm{s}^4(t)} \wedge \frac{\mathrm{w}^4(t)}{\mathrm{s}^2(t)n|\mathcal{M}_c[\varphi](t)|^2}\right)d\lambda(t) \\ &\quad + \int_{-k}^{k}|\mathcal{M}_c[f](t)|^2\,\mathrm{s}^2(t)\left(\frac{1}{n^2|\mathcal{M}_c[\varphi](t)|^4} \wedge \frac{1}{s^2(t)n|\mathcal{M}_c[\varphi](t)|^2}\right)\mathrm{w}^4(t)d\lambda(t) \\ &\leq L^2\left\|\frac{\mathrm{w}^4}{\mathrm{s}^4} \wedge \frac{\mathrm{w}^4}{\mathrm{s}^2\,n|\mathcal{M}_c[\varphi]|^2}\right\|_{\mathbb{L}^\infty} + \frac{L^2}{n^2}\mathbf{m}_k.\end{aligned}$$

Note that for $f \in \mathscr{F}(\mathrm{s}, L)$ we uniformly bound

$$\mathrm{v}^2_{f|\varphi} \leq (\|\varphi\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})}\|f\|_{\mathbb{L}^1_+(\mathrm{x}^{2c-2})} \vee 1) \leq (L^2\|\varphi\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})} \vee 1).$$

Consequently, using Proposition 3.2.1 we have that

$$\begin{aligned}&\sup_{f\in\mathscr{F}(\mathrm{s},L)} \mathbb{E}^n_g[|\hat{q}_k^2 - q^2|^2] \\ &\leq 2\mathrm{C}_\varphi \mathrm{v}^2_{f|\varphi}\left(L^4\frac{\mathrm{w}^4(k)}{\mathrm{s}^4(k)} + L^2\left\|\frac{\mathrm{w}^4}{\mathrm{s}^4} \wedge \frac{\mathrm{w}^4}{\mathrm{s}^2\,n|\mathcal{M}_c[\varphi]|^2}\right\|_{\mathbb{L}^\infty} + (L^2+1)\frac{\mathbf{m}_k \vee \mathbf{v}_k}{n^2}\right).\end{aligned}$$

The claim follows with $C = 6\mathrm{C}_\varphi(L^2\|\varphi\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})} \vee 1)(L^4 + 2L^2 + 1)$, which completes the proof. □

We illustrate the order of the upper bound in Corollary 3.3.2 setting

$$r^2(\mathscr{F}(\mathrm{s}, L)) := \inf_{k \in \mathbb{R}_+} \sup_{f \in \mathscr{F}(\mathrm{s},L)} \mathbb{E}_g^n[|\hat{q}_k^2 - q^2|^2] \tag{3.22}$$

under typical regularity assumptions. For this we define

$$k_* \in \arg \inf_{k \in \mathbb{R}_+} R_n(k) = \arg \inf_{k \in \mathbb{R}_+} \left(\frac{\mathrm{w}^4(k)}{\mathrm{s}^4(k)} \vee \frac{\mathbf{m}_k \vee \mathbf{v}_k}{n^2} \right). \tag{3.23}$$

By inserting k_* into the right-hand side of Equation (3.21) in Corollary 3.3.2 we derive the order of an upper bound of $r^2(\mathscr{F}(\mathrm{s}, L))$. As in Brenner Miguel et al. (2023) and Brenner Miguel et al. (2021), we consider the following examples. Concerning the class $\mathscr{F}(\mathrm{s}, L)$ defined in Equation (3.18) we distinguish the two behaviors of s described above. Namely the ordinary smooth case and the super smooth case. Regarding the error density φ, we distinguish again between two cases. We either assume for some decay parameter $\sigma \in \mathbb{R}_+$ its *(ordinary) smoothness*, i.e., $|\mathcal{M}_c[\varphi](t)| \sim (1+t^2)^{-\sigma/2}$; or its *super smoothness*, i.e., $|\mathcal{M}_c[\varphi](t)| \sim \exp(-|t|^\sigma)$. We restrict our discussion on the examples $\mathrm{w}(t) \sim t^a$ for $a \in \mathbb{R}$. Recall that in case of ordinary smoothness of the unknown density f Assumption 3.3.1 imposes the condition $s > a$ on the parameters. The order of the upper bound is given in Table 3.1 for the cases where both, the unknown density f and the error density φ are ordinarily smooth (first line), or one of them is ordinarily smooth and one is super smooth (line two and three).

$\mathrm{s}(t)$	$\lvert\mathcal{M}_c[\varphi](t)\rvert$	$R_n(k_*)$	R_n^{elbow}	$r^2(\mathscr{F}(\mathrm{s}, L))$
$(1+t^2)^{\frac{s}{2}}$	$(1+t^2)^{-\frac{\sigma}{2}}$	$n^{-\frac{8(s-a)}{4s+4\sigma+1}}$	$n^{-\frac{8(s-a)}{4s+4\sigma+1} \wedge 1}$	$\begin{cases} n^{-1}, & s \geq \sigma + 2a + \frac{1}{4} \\ n^{-\frac{8(s-a)}{4s+4\sigma+1}}, & s < \sigma + 2a + \frac{1}{4} \end{cases}$
$(1+t^2)^{\frac{s}{2}}$	$\exp(-\lvert t\rvert^\sigma)$	$(\log n)^{-\frac{4(s-a)}{\sigma}}$	$(\log n)^{-\frac{4(s-a)}{\sigma}}$	$(\log n)^{-\frac{4(s-a)}{\sigma}}$
$\exp(\lvert t\rvert^s)$	$(1+t^2)^{-\frac{\sigma}{2}}$	$\frac{1}{n^2}(\log n)^{\frac{4(\sigma+a)+1}{s}}$	n^{-1}	n^{-1}

Table 3.1: Order of the upper bound for $\mathrm{w}(t) \sim t^a$ for $t \in \mathbb{R},\ a \in \mathbb{R}$.

We omit the case that both f and φ are super smooth, since there are multiple possibilities how the rates behave, depending on the parameters. For a more detailed discussion of this case in the additive convolution model see for example Lacour (2006). The rates correspond to the rates derived by Butucea (2007) for the additive convolution model and Schluttenhofer and Johannes (2020b,a) for the circular convolution model. Both have shown that in the respective cases the rates are minimax. This suggests that this is also the case in the multiplicative convolution model. However, the proof of lower bounds is delayed to future work.

3.4 Data-driven estimation

The optimal choice $k_* \in \mathbb{R}_+$, see Equation (3.23), for estimator $\widehat{\theta}_{k_*}$, defined in Equation (3.4), depends on regularity parameter s of the unknown density f, which is not known in general. This motivates the consideration of data-driven procedures. The data-driven method is inspired by a bandwidth selection method in kernel density estimation proposed in Goldenshluger and Lepski

(2011). Inspired by Lepski's method (which appeared in a series of papers by Lepski (1990, 1991, 1992a,b)), given an upper bound $M \in \mathbb{N}$ and a sequence of penalties $(V(k))_{k\in\mathbb{N}}$, define the contrast

$$A(k) := \max_{k'\in[\![M]\!]} \left(|\hat{q}^2_{k\wedge k'} - \hat{q}^2_{k'}|^2 - V(k') - V(k)\right)_+, \tag{3.24}$$

where we write $a_+ := a \vee 0$ for $a \in \mathbb{R}$. In the spirit of Goldenshluger and Lepski (2011) combining the contrast given in Equation (3.24) and the penalization approach of model selection introduced by Barron et al. (1999) (for an extensive overview of model selection by penalized contrast the reader may refer to Massart (2007)) we select the dimension

$$\widetilde{k} := \arg\min_{k\in[\![M]\!]} \{A(k) + V(k)\}. \tag{3.25}$$

The data-driven estimator of q^2 is given by $\hat{q}^2_{\widetilde{k}}$. We derive an upper bound of its mean squared error $\mathbb{E}^n_g\left[|\hat{q}^2_{\widetilde{k}} - q^2|^2\right]$. This goal is achieved in two steps. First, we introduce in Section 3.4.1 a penalty series which still depends on the unknown error density f and derive an upper bound for the mean squared error of the resulting partially data-driven estimator. In a second step, we estimate the introduced penalty and therefore propose a fully data-driven estimator in Section 3.4.2 for which the upper bound of its mean squared error is derived based on the result for the partially data-driven estimator.

3.4.1 Partially data-driven penalty

In this section, we introduce a partially data-driven penalty. More precisely, for $k \in \mathbb{N}$ we set for some numerical constant $\kappa > 0$

$$\begin{aligned} V(k) := &\frac{\kappa \log(\Omega_k)}{n^2}\left(\mathbf{v}_k \vee (\log\Omega_k)\mathbf{m}_k\right) \\ &\cdot \left(\mathrm{C}^2_\varphi \mathrm{v}^2_{f|\varphi} + \left|1 \vee \frac{\Omega_k(\log\Omega_k)((2k)\vee(\log\Omega_k))}{n}\right|^2\right), \end{aligned} \tag{3.26}$$

where $\mathrm{v}_{f|\varphi} = (\|g\|_{\mathbb{L}^1_+(\mathrm{x}^{2c-2})} \vee 1)$, and

$$\Omega_k := 2k^3\left(1 \vee n^{-1}\mathbf{m}_k\right), \tag{3.27}$$

for $\mathbf{v}_k$ defined in Equation (3.7) and $\mathbf{m}_k$ in Equation (3.19). Note that $V(k)$ depends on $\mathrm{v}_{f|\varphi}$ and, therefore, on f. Consequently, it is unknown and needs to be estimated in a second step, see Section 3.4.2. In addition, $V(k)$ depends on the sample size n. However, for sake of simplicity we will omit additional subscripts. We denote by $\mathcal{C}$ a universal numerical constant which might change from line to line.

Theorem 3.4.1:
Under Assumption 3.1.1 the estimator $\hat{q}^2_{\widetilde{k}}$ given in Equation (3.4) with $\widetilde{k}$ defined in Equation (3.25) and arbitrary $M \in \mathbb{N}$, satisfies that there exists a universal numerical constant $\mathcal{C}$ such that

$$\begin{aligned} &\mathbb{E}^n_g\left[|\hat{q}^2_{\widetilde{k}} - q^2|^2\right] \\ &\le \mathcal{C} \min_{k\in[\![M]\!]}\left((q^2_k - q^2)^2 + V(k) + \frac{\mathrm{C}_\varphi \mathrm{v}^2_{f|\varphi}}{n}\Lambda_{f|\varphi}(k)\right) + \mathcal{C}\frac{\mathrm{C}^3_\varphi \mathrm{v}^3_{f|\varphi}}{n}(1 \vee (\mathbb{E}_g[\mathrm{Y}_1^{4(c-1)}])^2), \end{aligned}$$

for $\Lambda_{f|\varphi}(k)$ defined in Equation (3.7), $V(k)$ in Equation (3.26) with some numerical constant $\kappa > 0$.

We first discuss the result, give a data-driven choice for parameter M and give corresponding rates, before giving a proof of Theorem 3.4.1 at the end of this section.

Remark 3.4.2:
Let us compare the partially data-driven result Theorem 3.4.1 with Proposition 3.2.1. For this, we define

$$\rho_{k,n} := \log(\Omega_k)\left|1 \vee \frac{\Omega_k \log(\Omega_k)((2k) \vee \log(\Omega_k))}{n}\right|^2 \tag{3.28}$$

and

$$V_\varphi(k) := \rho_{k,n}\left(\mathbf{v}_k \vee \log(\Omega_k)\mathbf{m}_k\right). \tag{3.29}$$

Note that $V_\varphi(k)$ depends on φ but not on f. The rate in Theorem 3.4.1 is determined by

$$R_n^{\mathrm{pd}} := \min_{k\in\mathbb{R}_+}\left((q_k^2 - q^2)^2 + \frac{V_\varphi(k)}{n^2} + \frac{\Lambda_{f|\varphi}(k)}{n}\right).$$

Compared to the upper bound in Proposition 3.2.1, which is given by

$$\min_{k\in\mathbb{R}_+}\left((q_k^2 - q^2)^2 + \frac{\mathbf{v}_k}{n^2} + \frac{\Lambda_{f|\varphi}(k)}{n}\right),$$

the price we pay for a data-driven approach is on the one hand the additional factor $\rho_{k,n}$ of Equation (3.28) and on the other hand the term $\log(\Omega_k)\mathbf{m}_k$. We will see that the last term is often negligible with respect to $\mathbf{v}_k$. For details, refer to the discussion of the rates below. In addition, we minimize the parameter k over the set $[\![M]\!]$ instead of $\mathbb{R}_+$. If the minimum is attained in this set, there is no additional deterioration of the rate. The following result provides a data-driven choice for M.

Corollary 3.4.3 (Upper bound)**:**
Let Assumption 3.1.1 be satisfied and we use the notation of Remark 3.4.2. For $n \in \mathbb{N}$, set

$$M_\varphi^n := \max_{k\in\mathbb{N}}(V_\varphi(k) \le n^2 V_\varphi(1))$$

and $\widetilde{k}$ defined in Equation (3.25) choosing $M = M_\varphi^n$. Then, from Theorem 3.4.1 we immediately obtain that there exist constants $C_1, C_2 > 0$ such that for all $n \in \mathbb{N}$ with $V_\varphi(1) \ge R_n^{\mathrm{pd}}$ it holds that

$$\mathbb{E}_g^n\left[|\hat{q}_{\widetilde{k}}^2 - q^2|^2\right] \le C_1 R_n^{\mathrm{pd}} + \frac{C_2}{n}.$$

Proof of Corollary 3.4.3. Denote

$$k_n := \arg\min_{k\in\mathbb{N}}\left((q_k^2 - q^2)^2 + \frac{V_\varphi(k)}{n^2} + \frac{1}{n}\Lambda_{f|\varphi}(k)\right).$$

For $n \in \mathbb{N}$ with $V_\varphi(1) \ge R_n^{\mathrm{pd}}$ we have that

$$V_\varphi(1) \ge R_n^{\mathrm{pd}} \ge \frac{V_\varphi(k_n)}{n^2}$$

from which follows that $k_n \in [\![M_\varphi^n]\!]$ by definition of M_φ^n. With Assumption 3.1.1 (ii), we get that for the term $V(k)$ defined in Equation (3.26) there exists a constant $C \in \mathbb{R}_+$ such that

$$V(k_n) \leq C \frac{V_\varphi(k_n)}{n^2}.$$

With the calculations of Equation (3.31) follows the result. □

Define

$$R_n^{\text{fd}}(k) := \frac{\text{w}^4(k)}{\text{s}^4(k)} \vee \frac{\rho_{n,k}}{n^2} \left(\log(\Omega_k) \mathbf{m}_k \vee \mathbf{v}_k \right). \tag{3.30}$$

Analogously to Corollary 3.3.2 one can show the following result and we omit the proof.

Corollary 3.4.4:
Under Assumptions Assumption 3.1.1 and Assumption 3.3.1, if $f \in \mathscr{F}(\text{s}, L)$, see Equation (3.18), and $\mathbb{E}_g[Y_1^{4(c-1)}] \leq L$ then the estimator $\hat{q}^2_{\widetilde{k}}$ given in Equation (3.4) satisfies for $\widetilde{k}$ defined in Equation (3.25), $k \in [\![M]\!]$ and $n \in \mathbb{N}$, $n \geq 2$ that

$$\sup_{f \in \mathscr{F}(\text{s},L)} \mathbb{E}_g^n[|\hat{q}^2_{\widetilde{k}} - q^2|^2] \leq C \left(R_n^{\text{fd}}(k) \vee R_n^{\text{elbow}} \right)$$

for some constant $C > 0$ depending on L and φ, R_n^{elbow} in Equation (3.20) and $R_n^{\text{fd}}(k)$ in Equation (3.30).

Before illustrating the result let us give a few remarks.

Remark 3.4.5:
We see that the term R_n^{elbow} is already known from result Corollary 3.3.2. As already stated in Remark 3.4.2, we see that the price we pay for a data-driven approach is in the first term of the result, more precisely, in the terms $\rho_{n,k}$ and $\log(\Omega_k)\mathbf{m}_k$. Although the latter is often negligible with respect to $\mathbf{v}_k$.

Analogously to Equation (3.22) we define

$$R^2(\mathscr{F}(\text{s}, L)) := \inf_{k \in \mathbb{R}_+} \sup_{f \in \mathscr{F}(\text{s},L)} \mathbb{E}_g^n[|\hat{q}^2_k - q^2|^2]$$

and

$$k_{\text{opt}} \in \arg \inf_{k \in \mathbb{R}_+} R_n^{\text{fd}}(k) = \arg \inf_{k \in \mathbb{R}_+} \left(\frac{\text{w}^4(k)}{\text{s}^4(k)} \vee \frac{\rho_{n,k}}{n^2} \left(\log(\Omega_k) \mathbf{m}_k \vee \mathbf{v}_k \right) \right).$$

By inserting k_{opt} into the right-hand side of the bound in Corollary 3.4.4 we derive the order of an upper bound of $R^2(\mathscr{F}(\text{s}, L))$ for usual assumptions on the regularity as discussed in Section 3.3. More precisely, we give the order of $R_n^{\text{fd}}(k)$, R_n^{elbow} and $R^2(\mathscr{F}(\text{s}, L))$ below in Table 3.2 for each of the three cases already considered in Table 3.1. For this, we derive the order of $R_n^{\text{fd}}(k)$ for each of the examples and discuss its impact on the order of $R^2(\mathscr{F}(\text{s}, L))$.

(i) In the case when both the unknown density f and the error density φ are ordinarily smooth (first line of Table 3.2), we obtain different rates depending on the parameters σ, s and a. First, we see that for $k \in \mathbb{N}$

$$\Omega_k \lesssim k^3 \left(1 \vee \frac{1}{n} k^{4(a+\sigma)}\right) \le k^{3+4(a+\sigma)}$$

and, consequently, $\log(\Omega_k) \lesssim (1 \vee \log(k))$. In this case the term $\mathbf{v}_k$ is dominating $\log(\Omega_k)\mathbf{m}_k$ with rate $k^{4(a+\sigma)+1}$ and the bias-term is of order $k^{-4(s-a)}$, assuming that $a \le s$ and that $4(\sigma + a) > -1$. Consequently, to obtain an upper bound for the optimal choice of k we have to solve

$$\begin{aligned} 1 &\sim \frac{V_\varphi k}{n^2} \frac{\mathrm{s}^4(k)}{\mathrm{w}^4(k)} + \sim \left(1 \vee \frac{k^4(1 \vee \frac{1}{n} k^{4(a+\sigma)}) \log(k)}{n}\right)^2 \frac{\log(k) \vee 1}{n^2} k^{4(s+\sigma)+1} \\ &= \left(1 \vee \frac{k^4 \log(k)}{n} \vee \frac{k^{4+4(a+\sigma)}}{n^2}\right)^2 \frac{\log(k) \vee 1}{n^2} k^{4(s+\sigma)+1}. \end{aligned}$$

For simplicity, assume $(\sigma + 1) \ge 1$ and $(s - a) \ge \frac{3}{4}$. In this case, we observe again the elbow effect. More precisely, choosing $k_{\text{opt}} = \left(\frac{n^2}{\log(n)}\right)^{\frac{1}{4s+4\sigma+1}}$ in the case of $s < \sigma + 2a + \frac{1}{4}$ this term dominates the term R_n^{elbow}. On the other hand, in the case of $s \ge \sigma + 2a + \frac{1}{4}$ we retain the rate n^{-1}. Slower rates are achieved under other assumptions on the parameters s, σ and a.

(ii) Considering the case that the unknown density f is ordinarily smooth and the error density φ is super smooth (second line of Table 3.2), we see that the rate can be retained and the dimension parameter only changes slightly. First, $\mathbf{v}_k$ and $\log(\Omega_k)\mathbf{m}_k$ are both of order $k^{4a}\exp(4k^\sigma)$. The bias is of order $k^{-4(s-a)}$. Consequently, we need to find k for which holds

$$1 \sim \frac{\log(\Omega_k)}{n^2} k^{4s} \exp\left(4k^\sigma\right) \left|1 \vee \frac{\Omega_k \log(\Omega_k)((2k) \vee \log(\Omega_k))}{n}\right|^2.$$

For $k_{\text{opt}} \sim \left(\frac{\log n}{4} - \frac{4s}{\sigma} \log\left(\frac{\log n}{4}\right)\right)^{1/\sigma}$ we see that

$$\begin{aligned} \Omega_{k_{\text{opt}}} &\sim k_{\text{opt}}^3 (1 \vee n^{-1} k_{\text{opt}}^{4a} \exp(4k_{\text{opt}}^\sigma)) \\ &\sim \left(\frac{\log(n)}{4}\right)^{\frac{3}{\sigma}} \left(1 \vee \frac{1}{n}\left(\frac{\log(n)}{4}\right)^{\frac{4a}{\sigma}} \left(\frac{\log(n)}{4}\right)^{\frac{-4s}{\sigma}}\right) \sim \left(\frac{\log(n)}{4}\right)^{\frac{3}{\sigma}} \end{aligned}$$

and hence $\log(\Omega_{k_{\text{opt}}})$ is of order $\log(\log(n))$ fulfilling the above stated condition.

(iii) For the unknown density f being super smooth and the error density φ ordinarily smooth (third line of Table 3.2), we solve

$$1 \sim \frac{\log(\Omega_k)}{n^2} k^{4(a+\sigma)+1} k^{-4a} \exp\left(4\alpha k^s\right) \left|1 \vee \frac{\Omega_k \log(\Omega_k)((2k) \vee \log(\Omega_k))}{n}\right|^2.$$

For $k_{\text{opt}} \sim \left(\frac{\log n}{4} - \frac{4\sigma+1}{4s} \log\left(\frac{\log n}{4}\right)\right)^{1/s}$, $\Omega_{k_{\text{opt}}}$ and $\log(\Omega_{k_{\text{opt}}})$ behave as in part (ii) with rate $\log(n)$ and $\log(\log(n))$. Similarly to part (ii) the first term is of order 1. In case of $4(a + \sigma) + 1 < 0$ the rate n^{-1} is retained. Otherwise, there is a log-loss.

$\mathrm{s}(t)$	$\lvert\mathcal{M}_c[\varphi](t)\rvert$	$R_n^{\mathrm{fd}}(k_{\mathrm{opt}})$	R_n^{elbow}	$R^2(\mathscr{F}(\mathrm{s}, L))$
$(1+t^2)^{\frac{s}{2}}$	$(1+t^2)^{-\frac{\sigma}{2}}$	$\left(\frac{n^2}{\log n}\right)^{-\frac{4(s-a)}{4s+4\sigma+1}}$	$n^{-\frac{8(s-a)}{4s+4\sigma+1}\wedge 1}$	$\begin{cases} n^{-1}, & s \geq \sigma + 2a + \frac{1}{4} \\ \left(\frac{n^2}{\log n}\right)^{-\frac{4(s-a)}{4s+4\sigma+1}}, & s < \sigma + 2a + \frac{1}{4} \end{cases}$
$(1+t^2)^{\frac{s}{2}}$	$\exp(-\lvert t\rvert^{\sigma})$	$(\log n)^{-\frac{4(s-a)}{\sigma}}$	$(\log n)^{-\frac{4(s-a)}{\sigma}}$	$(\log n)^{-\frac{4(s-a)}{\sigma}}$
$\exp(\lvert t\rvert^{s})$	$(1+t^2)^{-\frac{\sigma}{2}}$	$\frac{(\log n)^{\frac{4(a-\sigma)+1}{s}}}{n}$	n^{-1}	$\frac{(\log n)^{\frac{4(a-\sigma)+1}{s}}}{n}$

Table 3.2: Upper bound of the order of the estimation risk for $\mathrm{w}(t) \sim t^a$ for $t \in \mathbb{R}$, $a \in \mathbb{R}$.

Finally, let us now turn to the proof of Theorem 3.4.1.

Proof of Theorem 3.4.1. Using that for $k, k' \in [\![M]\!]$ it holds

$$|\hat{q}_k^2 - \hat{q}_{k'}^2|^2 - V(k') - V(k) \leq A(k \wedge k') \leq A(k) + A(k'),$$

we have for any $k \in [\![M]\!]$ that

$$\begin{aligned}
|\hat{q}_{\widetilde{k}}^2 - q^2|^2 &\leq 2|\hat{q}_{\widetilde{k}}^2 - \hat{q}_k^2|^2 + 2|\hat{q}_k^2 - q^2|^2 \\
&\leq 2\left(|\hat{q}_{\widetilde{k}}^2 - \hat{q}_k^2|^2 - V(k) + V(k) - V(\widetilde{k}) + V(\widetilde{k})\right) + 2|\hat{q}_k^2 - q^2|^2 \\
&\leq 2(A(k) + A(\widetilde{k}) + V(k) + V(\widetilde{k})) + 2|\hat{q}_k^2 - q^2|^2 \\
&\leq 4A(k) + 4V(k) + 2|\hat{q}_k^2 - q^2|^2. \qquad (3.31)
\end{aligned}$$

Next, study $\mathbb{E}_g^n[A(k)]$ for any $k \in [\![M]\!]$. For this, we first decompose $A(k)$ reasonably. Using the decomposition Equation (3.13) we get that

$$\begin{aligned}
A(k) &= \max_{k < k' \leq M} \left(|\hat{q}_k^2 - \hat{q}_{k'}^2|^2 - V(k') - V(k)\right)_+ \\
&= \max_{k < k' \leq M} \left(|\hat{q}_k^2 - q_k^2 - (\hat{q}_{k'}^2 - q_{k'}^2) + q_k^2 - q_{k'}^2|^2 - V(k') - V(k)\right)_+ \\
&= \max_{k < k' \leq M} \left(|U_k - U_{k'} + 2(W_k - W_{k'}) + q_k^2 - q_{k'}^2|^2 - V(k') - V(k)\right)_+
\end{aligned}$$

Further, note that $(q_k^2 - q_{k'}^2)^2 \leq (q_{k \wedge k'}^2 - q^2)^2$. We subtract and add additionally $5(q_k^2 - q^2)^2$ which will be later useful to handle the linear part. With this we get

$$\begin{aligned}
&A(k) \\
&\leq \max_{k < k' \leq M} \left(4|U_k|^2 + 4|U_{k'}|^2 + 16\,|W_k - W_{k'}|^2 + 4(q_k^2 - q^2)^2 - V(k') - V(k)\right)_+ \\
&\leq \max_{k < k' \leq M} \left(4|U_k|^2 + 4|U_{k'}|^2 + 16\,|W_k - W_{k'}|^2\right. \\
&\qquad \left. - (V(k') + V(k) + (5-4)(q_k^2 - q^2)^2)\right)_+ + 5(q_k^2 - q^2)^2. \qquad (3.32)
\end{aligned}$$

Putting the calculations together and using Proposition 3.2.1 we get that

$$
\begin{aligned}
&\mathbb{E}_g^n\left[|\hat{q}_{\tilde{k}}^2 - q^2|^2\right] \\
&\leq 4\mathbb{E}_g^n\left[\max_{k<k'\leq M}\left(4|U_{k'}|^2 - \frac{1}{2}V(k')\right)_+\right] + 4\mathbb{E}_g^n\left[(4|U_k|^2 - V(k))_+\right] \\
&\quad + 4\mathbb{E}_g^n\left[\max_{k<k'\leq M}\left(16|W_{k'} - W_k|^2 - (\frac{1}{2}V(k') + (q_k^2 - q^2)^2)\right)_+\right] \\
&\quad + 20(q_k^2 - q^2)^2 + 2\mathbb{E}_g^n\left[|\hat{q}_k^2 - q^2|^2\right] + 4V(k) \\
&\leq 32\mathbb{E}_g^n\left[\max_{k\leq k'\leq M}\left(|U_{k'}|^2 - \frac{1}{8}V(k')\right)_+\right] \\
&\quad + 64\mathbb{E}_g^n\left[\max_{k<k'\leq M}\left(|W_{k'} - W_k|^2 - (\frac{1}{32}V(k') + \frac{1}{16}(q_k^2 - q^2)^2)\right)_+\right] \\
&\quad + 24(q_k^2 - q^2)^2 + 4\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}^2\left(\frac{\mathbf{v}_k}{n^2} + \frac{\Lambda_{f|\varphi}(k)}{n}\right) + 4V(k).
\end{aligned}
$$

Consequently, we split the contrast term $A(k)$, into a U-statistic part and a linear part. In the second step, we need to split each term again in a bounded and an unbounded term, see Lemma 3.7.1 and Lemma 3.7.5. We do this, to then apply an exponential inequality, in Lemma 3.7.2, and a Bernstein inequality, in Lemma 3.7.6. Lastly, we show that both unbounded parts are negligible, in Lemma 3.7.4 and Lemma 3.7.7. Choosing $\kappa \geq 16^2\kappa^*$ with κ^* as in concentration inequality Lemma 3.6.2 the assumptions of Lemma 3.7.1 and Lemma 3.7.5 are fulfilled. Thus, applying Lemma 3.7.1 and Lemma 3.7.5, and using that $V(k) \geq \mathrm{C}_\varphi \mathrm{v}_{f|\varphi}^2 \mathbf{v}_k n^{-2}$ yields the result. □

3.4.2 Fully data-driven penalty

Recall that in the definition of $V(k)$, see Equation (3.26), appears the constant $\mathrm{v}_{f|\varphi}^2$ defined in Equation (3.8) and consequently also in the definition of $\tilde{k}$, see Equation (3.25). The constant depends on the unknown density f. For this, first consider the estimator $\widehat{\mathrm{v}}^2 := 1 + \frac{1}{n}\sum_{j\in[\![n]\!]} Y_j^{4(c-1)}$ which is unbiased for the parameter $\mathrm{v}^2 := 1 + \mathbb{E}_g[Y_1^{4(c-1)}] \geq \mathrm{v}_{f|\varphi}^2$. Based on this estimator let us introduce a fully data-driven sequence of penalties $\widehat{V}(k)$ for $k \in \mathbb{N}$ given by

$$
\begin{aligned}
\widehat{V}(k) := &\frac{\kappa\log(\Omega_k)}{n^2}(\mathbf{v}_k \vee \log(\Omega_k)\mathbf{m}_k) \\
&\cdot\left(2\widehat{\mathrm{v}}^2\mathrm{C}_\varphi^2 + \left|1 \vee \frac{\Omega_k\log(\Omega_k)((2k)\vee\log(\Omega_k))}{n}\right|^2\right)
\end{aligned}
$$

which is now fully known in advance. Further, we use the fully known upper bound $M_\varphi^n \in \mathbb{N}$ defined in Corollary 3.4.3. Considering the fully data-driven estimator $\hat{q}_{\widehat{k}}^2$ defined in Equation (3.4) with dimension parameter selected by Goldenshluger and Lepski's method

$$
\begin{aligned}
\widehat{A}(k) &:= \max_{k'\in[\![M_\varphi^n]\!]}\left(|\hat{q}_{k\wedge k'}^2 - \hat{q}_{k'}^2|^2 - \widehat{V}(k') - \widehat{V}(k)\right)_+, \\
\widehat{k} &:= \arg\min_{k\in[\![M_\varphi^n]\!]}\{\widehat{A}(k) + \widehat{V}(k)\}.
\end{aligned}
\tag{3.33}
$$

We derive an upper bound for its mean squared error.

Theorem 3.4.6:
Let $\mathbb{E}_g[Y_1^{8(c-1)}] < \infty$. Under Assumption 3.1.1 the estimator $\hat{q}^2_{\widehat{k}}$ defined in Equation (3.4) with dimension $\widehat{k}$ defined in Equation (3.33) satisfies that there exist constant $C_{f,\varphi}$ and a universal numerical constant $\mathcal{C}$ such that

$$\mathbb{E}^n_g\left[|\hat{q}^2_{\widehat{k}} - q^2|^2\right] \leq \mathcal{C} \min_{k\in[\![M^n_\varphi]\!]} \left((q^2_k - q^2)^2 + V(k) + \frac{\mathrm{C}_\varphi \mathrm{v}^2_{f|\varphi}}{n}\Lambda_{f|\varphi}(k)\right) + C_{f,\varphi}\frac{1}{n},$$

for M^n_φ defined in Corollary 3.4.3, $V(k)$ in Equation (3.26) with some numerical constant $\kappa > 0$ large enough, and $\Lambda_{f|\varphi}(k)$ in Equation (3.7).

Proof of Theorem 3.4.6. The proof follows along the lines of the proof of Theorem 3.4.1. First, we note that for all $k \in [\![M^n_\varphi]\!]$ we have

$$\widehat{A}(k) \leq A(k) + 2 \max_{k\leq k'\leq M^n_\varphi}\{(V(k') - \widehat{V}(k'))_+\}$$

and, thus, similar to Equation (3.31) follows

$$|\hat{q}^2_{\widehat{k}} - q^2|^2 \leq 4A(k) + 4\widehat{V}(k) + 2|\hat{q}^2_k - q^2|^2 + 8 \max_{k\leq k'\leq M^n_\varphi}\{(V(k') - \widehat{V}(k'))_+\}$$

Using Equation (3.32) we get

$$\begin{aligned}
&\mathbb{E}^n_g\left[|\hat{q}^2_{\widehat{k}} - q^2|^2\right] \\
&\leq 32\mathbb{E}^n_g\left[\max_{k\leq k'\leq M^n_\varphi}\left(|U_{k'}|^2 - \frac{1}{8}V(k')\right)_+\right] \\
&\quad + 64\mathbb{E}^n_g\left[\max_{k<k'\leq M^n_\varphi}\left(|W_{k'} - W_k|^2 - (\frac{1}{32}V(k') + \frac{1}{16}(q^2_k - q^2)^2)\right)_+\right] \\
&\quad + 24(q^2_k - q^2)^2 + 4\mathrm{C}_\varphi \mathrm{v}^2_{f|\varphi}\left(\frac{\mathbf{v}_k}{n^2} + \frac{\Lambda_{f|\varphi}(k)}{n}\right) \\
&\quad + 4\mathbb{E}^n_g[\widehat{V}(k)] + 8\mathbb{E}^n_g[\max_{k\leq k'\leq M^n_\varphi}\{(V(k') - \widehat{V}(k'))_+\}].
\end{aligned}$$

The first two summands we again control with Lemma 3.7.1 and Lemma 3.7.5. Choosing parameter $\kappa \geq 16^2\kappa^*$ with κ^* as in concentration inequality Lemma 3.6.2, the assumptions of Lemma 3.7.1 and Lemma 3.7.5 are fulfilled. Thus, applying Lemma 3.7.1 and Lemma 3.7.5 we get

$$\begin{aligned}
&\mathbb{E}^n_g\left[|\hat{q}^2_{\widehat{k}} - q^2|^2\right] \\
&\leq \mathcal{C}(q^2_k - q^2)^2 + 4\mathrm{C}_\varphi \mathrm{v}^2_{f|\varphi}\left(\frac{\mathbf{v}_k}{n^2} + \frac{\Lambda_{f|\varphi}(k)}{n}\right) + \mathcal{C}\frac{\mathrm{C}^3_\varphi \mathrm{v}^3_{f|\varphi}}{n}(1 \vee (\mathbb{E}_g[Y_1^{4(c-1)}])^2) \\
&\quad + 4\mathbb{E}^n_g[\widehat{V}(k)] + 8\mathbb{E}^n_g[\max_{k\leq k'\leq M^n_\varphi}\{(V(k') - \widehat{V}(k'))_+\}].
\end{aligned}$$

Moreover, we have $\mathbb{E}^n_g[\widehat{V}(k)] \leq 4V(k)$ and using that $V(k) \geq \mathrm{C}_\varphi \mathrm{v}^2_{f|\varphi}\mathbf{v}_k n^{-2}$ yields

$$\begin{aligned}
&\mathbb{E}^n_g\left[|\hat{q}^2_{\widehat{k}} - q^2|^2\right] \\
&\leq \mathcal{C}(q^2_k - q^2)^2 + \mathcal{C}\left(V(k) + \mathrm{C}_\varphi \mathrm{v}^2_{f|\varphi}\frac{\Lambda_{f|\varphi}(k)}{n}\right) + \mathcal{C}\frac{\mathrm{C}^3_\varphi \mathrm{v}^3_{f|\varphi}}{n}(1 \vee (\mathbb{E}_g[Y_1^{4(c-1)}])^2) \\
&\quad + 8\mathbb{E}^n_g[\max_{k\leq k'\leq M^n_\varphi}\{(V(k') - \widehat{V}(k'))_+\}]. \qquad (3.34)
\end{aligned}$$

Making use of $V_\varphi(k)$ defined in Equation (3.29) and $\max_{k\le k'\le M_\varphi^n} V_\varphi(k') \le n^2 V_\varphi(1)$ by the definition of M_φ^n we obtain

$$\begin{aligned}
&\max_{k\le k'\le M_\varphi^n}\{(V(k')-\widehat{V}(k'))_+\}\\
&\le \mathrm{C}_\varphi^2\kappa\left(\mathbb{E}_g[Y_1^{4(c-1)}]-\frac{2}{n}\sum_{j\in[\![n]\!]}Y_j^{4(c-1)}\right)_+\frac{1}{n^2}\max_{k\le k'\le M_\varphi^n}V_\varphi(k')\\
&\le \mathrm{C}_\varphi^2\kappa V_\varphi(1)\left(\mathbb{E}_g[Y_1^{4(c-1)}]-\frac{2}{n}\sum_{j\in[\![n]\!]}Y_j^{4(c-1)}\right)_+ .
\end{aligned}$$

For $a>0$ and $b\ge 0$ it holds that $(\frac{a}{2}-b)_+\le 2\frac{|a-b|^2}{a}$. Therefore, it follows

$$\begin{aligned}
\mathbb{E}_g^n\left[\max_{k\le k'\le M_\varphi^n}\{(V(k')-\widehat{V}(k'))_+\}\right] &\le \mathrm{C}_\varphi^2\kappa V_\varphi(1)\mathbb{E}_g^n[\left(\mathbb{E}_g[Y_1^{4(c-1)}]-\frac{2}{n}\sum_{j\in[\![n]\!]}Y_j^{4(c-1)}\right)_+]\\
&\le \frac{\mathcal{C}}{n}\frac{\mathrm{C}_\varphi^2V_\varphi(1)\mathbb{E}_g[Y_1^{8(c-1)}]}{\mathbb{E}_g[Y_1^{4(c-1)}]}.
\end{aligned}$$

The last bound together with Equation (3.34) and setting

$$C_{f,\varphi}=\mathcal{C}\mathrm{C}_\varphi^3\mathrm{v}_{f|\varphi}^3(1\vee(\mathbb{E}_g[Y_1^{4(c-1)}])^2+\frac{V_\varphi(1)\mathbb{E}_g[Y_1^{8(c-1)}]}{\mathbb{E}_g[Y_1^{4(c-1)}]})$$

concludes the proof. □

Remark 3.4.7:
It is remarkable that the upper bounds in Theorem 3.4.1 and Theorem 3.4.6 only differ in the constants, i.e $C_{f,\varphi}$ compared to $\mathcal{C}\mathrm{C}_\varphi^3\mathrm{v}_{f|\varphi}^3(1\vee(\mathbb{E}_g[Y_1^{4(c-1)}])^2)$. However, in Theorem 3.4.6 we impose only a higher moment assumption, i.e $\mathbb{E}_g[Y_1^{8(c-1)}]<\infty$ compared to $\mathbb{E}_g[Y_1^{4(c-1)}]<\infty$. Consequently, for the fully-data driven case we obtain under the assumptions of Corollary 3.4.4 that

$$\sup_{f\in\mathscr{F}(\mathrm{s},L)}\mathbb{E}_g^n[|\hat{q}_{\widehat{k}}^2-q^2|^2]\le C\left(R_n^{\mathrm{fd}}(k)\vee R_n^{\mathrm{elbow}}\right)$$

and the same rates as for the partially data-driven case, see Table 3.2.

3.5 Simulation study

In this section, we illustrate the behavior of the estimator $\hat{q}_k^2$ presented in Equation (3.4) and the fully data-driven choice $\widehat{k}$ given in Equation (3.33). To do so we consider the following example continuing with Section 1.4.3. That is,we set $c=0.5$ and $\mathrm{w}(t)=1$ for $t\in\mathbb{R}$. Recall that in this case we have $q^2=\|f\|_{\mathbb{L}_+^2}^2$. For illustration purposes, we consider one example of an ordinarily smooth density, where X is Beta(a,b)-distributed with parameters $a=2,b=1$, and one example of a super smooth density, where X is log-normally distributed with parameters $\mu = 0, \sigma^2 = 1$. More precisely,

$$f_1(x) = 2x\mathbb{1}_{(0,1)}(x),$$
$$f_2(x) = \frac{1}{\sqrt{2\pi x^2}} \exp(-\log(x)^2/2).$$

For error densities, we consider the following two examples of U being either Pareto-distributed or log-normally distributed, i.e.

$$\varphi_1(x) = \mathbb{1}_{(1,\infty)}(x)x^{-2},$$
$$\varphi_2(x) = \frac{1}{\sqrt{2\pi x^2}} \exp(-\log(x)^2/2).$$

The corresponding Mellin-transforms are given by

$$\mathcal{M}_{\frac{1}{2}}[f_1](t) = 2/(1.5 + 2\pi i t),$$
$$\mathcal{M}_{\frac{1}{2}}[f_2](t) = \exp\left((-0.5 + 2\pi i t)^2/2\right),$$
$$\mathcal{M}_{\frac{1}{2}}[\varphi_1](t) = (1.5 - 2\pi i t)^{-1},$$
$$\mathcal{M}_{\frac{1}{2}}[\varphi_2](t) = \exp\left((-0.5 + 2\pi i t)^2/2\right).$$

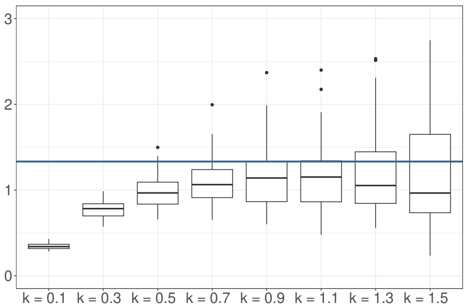
For a detailed discussion on the Mellin transforms and their decaying behavior of common probability densities, see Brenner Miguel (2023). We illustrate our results by considering the following three cases.

(i) Ordinary Smooth - Ordinary Smooth (OO): For $f = f_1$ and $\varphi = \varphi_1$ the true parameter is given by $q^2 = \|f_1\|_{\mathbb{L}^2_+} = 4/3 \approx 1.33$.

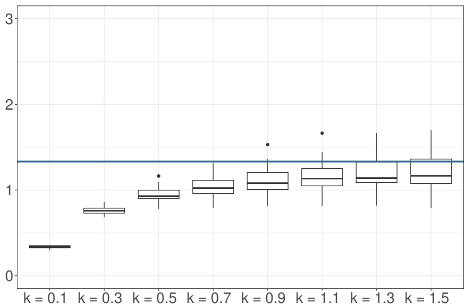
(ii) Super Smooth - Ordinary Smooth (SO): For $f = f_2$ and $\varphi = \varphi_1$ the true parameter is given by $q^2 = \|f_2\|_{\mathbb{L}^2_+} = \sqrt[4]{\exp(1)}/(2\sqrt{\pi}) \approx 0.3622$.

(iii) Ordinary Smooth - Super Smooth (OS): For $f = f_1$ and $\varphi = \varphi_2$ the true parameter is again given by $q^2 = 4/3 \approx 1.33$.

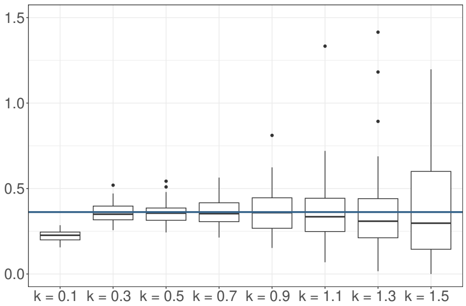
In Figure 3.1, for all three examples boxplots of the values of $\hat{q}^2_k$ are depicted over 50 iterations for $k \in \{0.1, 0.3, \ldots, 1.5\}$ and $n \in \{100, 500\}$. Whenever the values where negative, the estimator is set to zero. Consequently, if there is a large fluctuation, they are largely estimated to be zero. We note that the simulation results in these three cases reflect their corresponding rates of convergence $n^{-4/7}$, n^{-1} and $(\log n)^{-1}$ and highlight the importance of a proper choice of the dimension parameter k. Figure 3.2 shows for the same examples the fully data-driven estimates $\widehat{k}$, see Equation (3.33), with $\kappa = 0.00001$ over the set $k \in \{0.1, 0.2, \ldots, 2\}$. Here, we stopped the data-driven procedure as soon as there is too much fluctuation, i.e., as soon as there appeared the first negative estimated value. The results indicate that with increasing n the data-driven choice of k increases also, which is expected from the theoretical results. Further, for exponential decay in the error density results deteriorate.



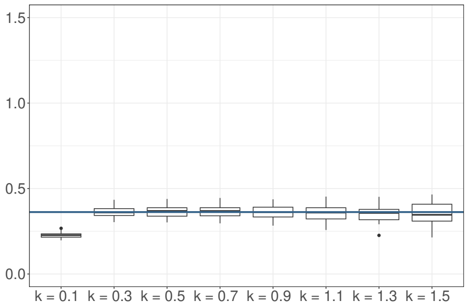
(a) Example (OO), $n = 100$



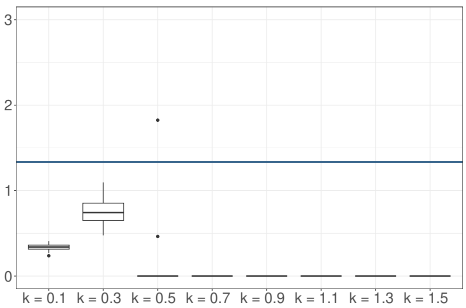
(b) Example (OO), $n = 500$



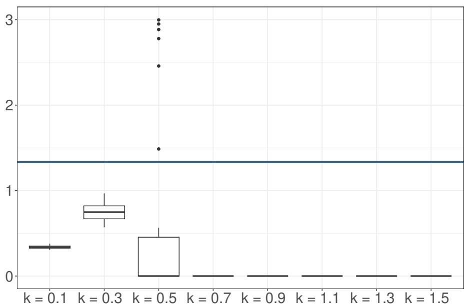
(c) Example (SO), $n = 100$



(d) Example (SO), $n = 500$



(e) Example (OS), $n = 100$



(f) Example (OS), $n = 500$

Figure 3.1: The boxplots represent the values of $\hat{q}_k^2$ for Examples (OO), (SO) and (OS) over 50 iterations. The horizontal lines indicate q^2.

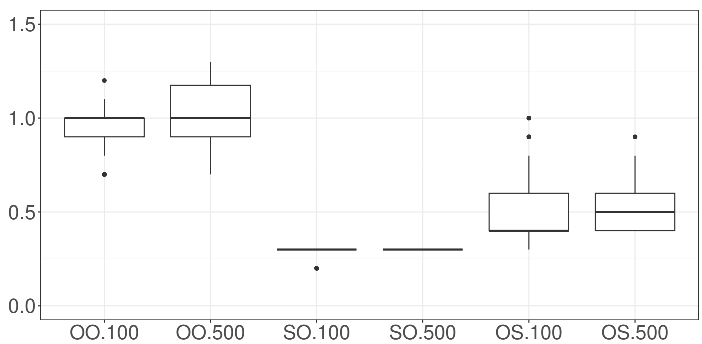


Figure 3.2: Values of the fully data-driven $\widehat{k}$ for Examples (OO), (SO) and (OS) over 50 iterations for $n \in \{100, 500\}$.

3.6 Auxiliary results

The next three assertions, a variance bound and a concentration inequality for canonical U-statistics and a Bernstein inequality, provide our key arguments in order to upper bound the risk of the quadratic functional estimator proposed in this chapter.

For $(X_j)_{j\in[\![n]\!]}$, $n \in \mathbb{N}$ independent and identically distributed random variables in $(\mathbb{R}_+, \mathscr{B}_+)$ and bounded symmetric kernel $h\colon \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$, a U-statistic is a sum of the form

$$U_n = \frac{1}{n(n-1)} \sum_{\substack{j_1 \neq j_2 \\ j_1, j_2 \in [\![n]\!]}} h(X_{j_1}, X_{j_2}).$$

The U-statistic is called canonical if for all $i, j \in [\![n]\!]$ and all $x, y \in \mathbb{R}_+$

$$\mathbb{E}\left[h(X_i, y)\right] = \mathbb{E}\left[h(x, X_j)\right] = 0.$$

We use the following adaption of Theorem 3, Chap.1, from Lee (1990).

Lemma 3.6.1:
Let $U := \frac{1}{n(n-1)} \sum_{\substack{j\neq l \\ j,l \in [\![n]\!]}} h(X_j, X_l)$ be a U-statistic with i.i.d. random variables $(X_j)_{j\in[\![n]\!]}$. Then

$$\mathrm{Var}(U) \leq \frac{1}{n(n-1)} \left(2(n-2)\mathbb{E}[|\mathbb{E}[h(X_1, X_2)]|^2] + \mathbb{E}^2[|h(X_1, X_2)|^2]\right).$$

The first bound in the next assertion, a concentration for canonical U-statistics, is a reformulation of Theorem 3.4.8 in Giné and Nickl (2016). For this lemma we have used the notation $\|\cdot\|_{\mathbb{L}^\infty_{+,2}}$ to indicate the essential supremum for functions on $\mathbb{R}_+ \times \mathbb{R}_+$.

Lemma 3.6.2:
Let U_n be a canonical U-statistic for $(X_j)_{j\in[\![n]\!]}$, $n \geq 2$, i.i.d. $\mathbb{R}_+$-valued random variables and kernel $h\colon \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ bounded and symmetric. Let

$$\begin{aligned}
&A \geq \|h\|_{\mathbb{L}^\infty_{+,2}}, \\
&B^2 \geq \|\mathbb{E}[h^2(X_1, \cdot)]\|_{\mathbb{L}^\infty_{+}}, \\
&C^2 \geq \mathbb{E}^2[h^2(X_1, X_2)], \\
&D \geq \sup\{\mathbb{E}^2[h(X_1, X_2)\xi(X_1)\zeta(X_2)] : \mathbb{E}[\xi^2(X_1)] \leq 1, \mathbb{E}[\zeta^2(X_2)] \leq 1\}.
\end{aligned}$$

Then, for all $x \geq 0$ it holds

$$\mathbb{P}^n\left(U_n \geq 8\frac{C}{n}x^{1/2} + 13\frac{D}{n}x + 261\frac{B}{n^{3/2}}x^{\frac{3}{2}} + 343\frac{A}{n^2}x^2\right) \leq \exp(1-x).$$

Hence, there exist universal numerical constants $\mathcal{C}, \kappa^*$ and d such that it also holds for any $K \geq 1$

$$\mathbb{E}^n\left[\left(|U_n|^2 - (\log K)\left(\frac{\kappa^* C^2}{2n^2} + \frac{\kappa^*(\log K)D^2}{32n^2} + \frac{d(\log K)^2 B^2}{4n^3} + \frac{d^2(\log K)^3 A^2}{16n^4}\right)\right)_+\right]$$
$$\leq \frac{\mathcal{C}}{K}\left(\frac{C^2}{n^2} + \frac{D^2}{n^2} + \frac{B^2}{n^3} + \frac{A^2}{n^4}\right).$$

Proof of Lemma 3.6.2. From the first bound, we obtain the second by integrating and choosing parameters $\kappa^* \geq 173.056$ and $d \geq 4.359.744$. □

The Bernstein inequality in the first formulation is for example given in Comte (2017), Appendix B, Lemma B.2. From the first bound, we obtain again the second by integrating.

Lemma 3.6.3 (Bernstein inequality)**:**
Let $(Z_j)_{j\in[\![n]\!]}$ be independent random variables satisfying $\mathbb{E}[Z_j] = 0$, $|Z_j| \leq b$ almost surely and $\mathbb{E}[|Z_j|^2] \leq v$ for all $j \in [\![n]\!]$. Then, for all $x > 0$ and $n \geq 1$, we have

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{j\in[\![n]\!]} Z_j\right| \geq x\right) \leq 2\max\left(\exp\left(\frac{-nx^2}{4v}\right), \exp\left(\frac{-nx}{4b}\right)\right).$$

Moreover, for any $K \geq 1$ we have

$$\mathbb{E}\left(\left(\Big|\frac{1}{\sqrt{n}}\sum_{j\in[\![n]\!]} Z_j\Big|^2 - (4v + 32b^2\log(K)n^{-1})\log(K)\right)_+\right) \leq \frac{8(v + 16b^2n^{-1})}{K}.$$

3.7 Proofs of Section 3.4

3.7.1 U-statistic results

Lemma 3.7.1 (Concentration of the U-statistic)**:**
Under Assumption 3.1.1 there exists a universal numerical constant $\mathcal{C} > 0$ such that

$$\mathbb{E}_g^n\left[\max_{k\in[\![M]\!]}\left(|U_k|^2 - \frac{1}{8}V(k)\right)_+\right] \leq \mathcal{C}C_\varphi^3 \mathrm{v}_{f|\varphi}^3(1 \vee \mathbb{E}_g[Y_1^{4(c-1)}])\frac{1}{n},$$

for U_k defined in Equation (3.11) and $V(k)$ in Equation (3.26) and $\kappa \geq 16^2\kappa^*$ for κ^* from Lemma 3.6.2.

Proof of Lemma 3.7.1. For $k \in \mathbb{N}$ consider the canonical U-statistic U_k defined in Equation (3.11) admitting as kernel the symmetric and real-valued function h_k given in Equation (3.15) (see proof of Proposition 3.2.1). We apply the concentration inequality for canonical U-statistics given in Lemma 3.6.2. Note that $|x^{c-1+2\pi it}| = x^{c-1}$ is not bounded for $x \in \mathbb{R}_+$ and hence $|h_k(x,y)|$ is generally not bounded for $x, y \in \mathbb{R}_+$. Therefore, we decompose h_k in bounded and a remaining unbounded part. More precisely, given $\delta_k \in \mathbb{R}_+$, we denote for $y \in \mathbb{R}_+$ and $t \in \mathbb{R}$

$$\psi_k^b(y,t) := \mathbb{1}_{[0,\delta_k]}(y^{c-1})y^{c-1+2\pi it} \quad\text{and}\quad \psi_k^u(y,t) := \mathbb{1}_{(\delta_k,\infty)}(y^{c-1})y^{c-1+2\pi it}. \tag{3.35}$$

Define the bounded part of kernel h_k as

$$h_k^b(x,y) := \int_{-k}^{k} \frac{(\psi_k^b(y,t) - \mathbb{E}_g[\psi_k^b(Y_1,t)])(\psi_k^b(x,-t) - \mathbb{E}_g[\psi_k^b(Y_1,-t)])}{|\mathcal{M}_c[\varphi](t)|^2}\mathrm{w}^2(t)d\lambda(t). \tag{3.36}$$

Then, h_k^b is indeed bounded since $|\psi_k^b(y,t)| \leq \delta_k$. Analogously, define

$$h_k^u(x,y) := \int_{-k}^{k} \frac{(\psi_k^u(y,t) - \mathbb{E}_g[\psi_k^u(Y_1,t)])(\psi_k^u(x,-t) - \mathbb{E}_g[\psi_k^u(Y_1,-t)])}{|\mathcal{M}_c[\varphi](t)|^2}\mathrm{w}^2(t)d\lambda(t) \tag{3.37}$$

and

$$h_k^{bu}(x,y) := \int_{-k}^{k} \frac{(\psi_k^u(y,t) - \mathbb{E}_g[\psi_k^u(Y_1,t)])(\psi_k^b(x,-t) - \mathbb{E}_g[\psi_k^b(Y_1,-t)])}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t)$$
$$+ \int_{-k}^{k} \frac{(\psi_k^b(y,t) - \mathbb{E}_g[\psi_k^b(Y_1,t)])(\psi_k^u(x,-t) - \mathbb{E}_g[\psi_k^u(Y_1,-t)])}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t). \quad (3.38)$$

Then, h_k^b, h_k^u and h_k^{bu} are also symmetric and real-valued and we have $h_k = h_k^b + h_k^u + h_k^{bu}$. Denote by U_k^b, U_k^u and U_k^{bu} the corresponding canonical U-statistics, respectively. Then, it holds that $U_k = U_k^b + U_k^u + U_k^{bu}$. Consequently, we obtain

$$\mathbb{E}_g^n\left[\max_{k\in[\![M]\!]}\left(|U_k|^2 - \frac{1}{8}V(k)\right)_+\right]$$
$$\leq 2\sum_{k\in[\![M]\!]}\mathbb{E}_g^n\left[\left(|U_k^b|^2 - \frac{1}{16}V(k)\right)_+\right] + 4\sum_{k\in[\![M]\!]}\left(\mathbb{E}_g^n[|U_k^u|^2] + \mathbb{E}_g^n[|U_k^{bu}|^2]\right). \quad (3.39)$$

We begin by considering the first summand. In Lemma 3.7.2 we apply the exponential inequality Lemma 3.6.2 to U_k^b. To be more precise, with Ω_k defined in Equation (3.27) and constants $\kappa^*, d > 0$ as in Lemma 3.6.2 we introduce

$$\delta_k^2 := \frac{\Omega_k}{32d\mathrm{C}_\varphi^2\mathrm{v}_{f|\varphi}^2} \quad \text{and} \quad K := \Omega_k^4, \quad (3.40)$$

and

$$V_1^b(k) := \log(K)\mathrm{C}_\varphi^2\mathrm{v}_{f|\varphi}^2\frac{1}{n^2}\left(\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^4(\mathrm{w}^4)} \vee \log(K)\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^\infty(\mathrm{w})}\right)$$
$$\cdot\left(\kappa^* + d\frac{\delta_k^2}{n}(\log(K))^2 + d^2\frac{\delta_k^4(2k)}{n^2}(\log(K))^3\right).$$

We show below in Equation (3.43) that $V_1^b(k) \leq \frac{1}{16}V(k)$. With this we get from Lemma 3.7.2 that

$$\mathbb{E}_g^n\left[\left(|U_k^b|^2 - \frac{1}{16}V(k)\right)_+\right] \leq \mathbb{E}_g^n\left[\left(|U_k^b|^2 - V_1^b(k)\right)_+\right]$$
$$\leq \frac{\mathcal{C}\mathrm{C}_\varphi^2\mathrm{v}_{f|\varphi}^2}{K}\frac{1}{n^2}\left(\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^4(\mathrm{w}^4)} \vee \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^\infty(\mathrm{w})}\right)$$
$$\cdot\left(1 + \frac{\delta_k^2}{n} + \frac{\delta_k^4(2k)}{n^2}\right).$$

Since by definition of Ω_k Equation (3.27) we have

$$\frac{1}{n}\left(\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^4(\mathrm{w}^4)} \vee \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^\infty(\mathrm{w})}\right) \leq \frac{\Omega_k}{k^2}, \quad (3.41)$$

$\Omega_k \geq 1$, by Equation (3.40) $\frac{\delta_k^2}{\Omega_k} \leq 1$ and $\frac{2k}{\Omega_k} \leq 1$ and

$$\left(1 + \frac{\delta_k^2}{n} + \frac{\delta_k^4(2k)}{n^2}\right) \leq 3\Omega_k^3.$$

Combining the last inequalities, Equation (3.40) and $\sum_{k\in\mathbb{N}} k^{-2} = \pi^2/6$ we get

$$\sum_{k\in[\![M]\!]} \mathbb{E}_g^n \left[\left(|U_k^b|^2 - \frac{1}{16} V(k) \right)_+ \right] \leq \sum_{k\in[\![M]\!]} \frac{\mathcal{C} \mathrm{C}_\varphi^2 \mathrm{v}_{f|\varphi}^2 \Omega_k^4}{K k^2} \frac{1}{n} \leq \frac{\mathcal{C} \mathrm{C}_\varphi^2 \mathrm{v}_{f|\varphi}^2}{n}. \tag{3.42}$$

To conclude the calculations for the first summand of Equation (3.39), we show now that it holds $V_1^b(k) \leq \frac{1}{16} V(k)$. Note that $K \geq 1$ for all $k \in \mathbb{N}$ and δ_k is increasing for $k \to \infty$. Further, we get

$$\begin{aligned}
&\mathrm{C}_\varphi^2 \mathrm{v}_{f|\varphi}^2 \left(\kappa^* + d\frac{\delta_k^2}{n} (\log(K))^2 + d^2 \frac{\delta_k^4(2k)}{n^2} (\log(K))^3 \right) \\
&\leq \kappa^* \mathrm{C}_\varphi^2 \mathrm{v}_{f|\varphi}^2 + \frac{\Omega_k (\log(\Omega_k))^2}{2n} + \frac{\Omega_k^2(2k)(\log(\Omega_k))^3}{2n^2} \\
&\leq \kappa^* \mathrm{C}_\varphi^2 \mathrm{v}_{f|\varphi}^2 + \left| 1 \vee \frac{\Omega_k \log(\Omega_k)((2k) \vee \log(\Omega_k))}{n} \right|^2.
\end{aligned}$$

Since $\kappa \geq 16^2 \kappa^*$ it follows

$$\begin{aligned}
V_1^b(k) &\leq 4\log(\Omega_k) \frac{1}{n^2} \left(\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^4(\mathrm{w}^4)}^4 \vee 4\log(\Omega_k) \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^\infty(\mathrm{w})}^4 \right) \\
&\quad \cdot \left(\kappa^* \mathrm{C}_\varphi^2 \mathrm{v}_{f|\varphi}^2 + \left| 1 \vee \frac{\Omega_k \log(\Omega_k)((2k) \vee \log(\Omega_k))}{n} \right|^2 \right) \\
&\leq 16\kappa^*/\kappa V(k) \leq \frac{1}{16} V(k).
\end{aligned} \tag{3.43}$$

Consider now the second summand of Equation (3.39) which we bound with the help of Lemma 3.7.4 below. Its proof is based on the variance bound for U-statistics Lemma 3.6.1 similarly as in Proposition 3.2.1. From Lemma 3.7.4 it follows

$$\sum_{k\in[\![M]\!]} (\mathbb{E}_g^n[|U_k^u|^2] + \mathbb{E}_g^n[|U_k^{bu}|^2]) \leq \sum_{k\in[\![M]\!]} \frac{10\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}}{n^2} \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^4(\mathrm{w}^4)}^4 \delta_k^{-2} \mathbb{E}_g[Y_1^{4(c-1)}].$$

Combining the last bound, Equation (3.41) and Equation (3.40) it follows that there exists a universal numerical constant $\mathcal{C}$ such that

$$\begin{aligned}
\sum_{k\in[\![M]\!]} (\mathbb{E}_g^n[|U_k^u|^2] + \mathbb{E}_g^n[|U_k^{bu}|^2]) &\leq 10\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} \mathbb{E}_g[Y_1^{4(c-1)}] \frac{1}{n} \sum_{k\in[\![M]\!]} \frac{\Omega_k}{k^2 \delta_k^2} \\
&\leq \mathcal{C} \mathrm{C}_\varphi^3 \mathrm{v}_{f|\varphi}^3 \mathbb{E}_g[Y_1^{4(c-1)}] \frac{1}{n} \sum_{k\in\mathbb{N}} \frac{1}{k^2} \leq \mathcal{C} \mathrm{C}_\varphi^3 \mathrm{v}_{f|\varphi}^3 \mathbb{E}_g[Y_1^{4(c-1)}] \frac{1}{n}.
\end{aligned}$$

Inserting the last bound and Equation (3.42) in Equation (3.39) implies the result. □

Lemma 3.7.2 (Concentration of the bounded U-statistic)**:**
For $k \in \mathbb{N}$ and $\delta_k \in \mathbb{R}_+$ the real-valued, bounded and symmetric kernel $h_k^b\colon \mathbb{R}^2 \to \mathbb{R}$ given in Equation (3.36) fulfills $\mathbb{E}_g[h_k^b(Y_1, y)] = 0$ for all $y \in \mathbb{R}_+$. Under Assumption 3.1.1 for the canonical U-statistic

$$U_k^b := \frac{1}{n(n-1)} \sum_{\substack{j \neq l \\ j,l \in [\![n]\!]}} h_k^b(Y_j, Y_l)$$

there exists a universal numerical constant $\mathcal{C}$ such that for each $K \geq 1$

$$\begin{aligned}
&\mathbb{E}_g^n\left[\left(|U_k^b|^2 - V_1^b(k)\right)_+\right] \\
&\leq \frac{\mathcal{C} C_\varphi^2 v_{f|\varphi}^2}{K} \frac{1}{n^2} \left(\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^4(w^4)}^4 \vee \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^\infty(w)}^4\right) \\
&\quad \cdot \left(1 + \frac{\delta_k^2}{n} + \frac{\delta_k^4(2k)}{n^2}\right),
\end{aligned}$$

with C_φ, $v_{f|\varphi}$ defined in Equation (3.8), numerical constants $\kappa^*, d > 0$ given in Lemma 3.6.2 and

$$\begin{aligned}
V_1^b(k) :=& (\log K) C_\varphi^2 v_{f|\varphi}^2 \frac{1}{n^2} \left(\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^4(w^4)}^4 \vee (\log K)\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^\infty(w)}^4\right) \\
&\cdot \left(\kappa^* + d\frac{\delta_k^2}{n}(\log K)^2 + d^2\frac{\delta_k^4(2k)}{n^2}(\log K)^3\right).
\end{aligned}$$

Proof of Lemma 3.7.2. We intend to apply the concentration inequality for canonical U-statistics given in Lemma 3.6.2 using the constants κ^* and d and the quantities A, B, C and D computed in Lemma 3.7.3. Using that $\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^2(w^2)}^4 \leq (2k)\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^\infty(w^2)}^4$ we get

$$\begin{aligned}
&\frac{\kappa^* C^2}{2n^2} + \frac{\kappa^*(\log K)D^2}{8n^2} + \frac{d(\log K)^2 B^2}{4n^3} + \frac{d^2(\log K)^3 A^2}{16n^4} \\
&\leq \kappa^* \frac{C_\varphi^2 v_{f|\varphi}^2}{n^2}\left(\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^4(w^4)}^4 \vee (\log K)\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^\infty(w)}^4\right) \\
&\quad + d\frac{(\log K)^2 C_\varphi v_{f|\varphi}\delta_k^2}{n^3}\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^4(w^4)}^4 \\
&\quad\quad + d^2\frac{(\log K)^3\delta_k^4}{n^4}\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^2(w^2)}^4 \\
&\leq \frac{C_\varphi^2 v_{f|\varphi}^2}{n^2}\left(\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^4(w^4)}^4 \vee (\log K)\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^\infty(w)}^4\right) \\
&\quad \cdot \left(\kappa^* + d\frac{\delta_k^2}{n}(\log K)^2 + d^2\frac{\delta_k^4(2k)}{n^2}(\log K)^3\right).
\end{aligned}$$

Consequently by Lemma 3.6.2 there exists a universal numerical constant $\mathcal{C}$ such that

$$\begin{aligned}
&\mathbb{E}_g^n\left[\left(|U_k^b|^2 - V_1^b(k)\right)_+\right] \\
&\leq \mathbb{E}_g^n\left[\left(|U_k^b|^2 - (\log K)\left(\frac{\kappa^* C^2}{2n^2} + \frac{\kappa^*(\log K)D^2}{8n^2} + \frac{d(\log K)^2 B^2}{4n^3} + \frac{d^2(\log K)^3 A^2}{16n^4}\right)\right)_+\right] \\
&\leq \frac{\mathcal{C}}{K}\left(\frac{C^2}{n^2} + \frac{D^2}{n^2} + \frac{B^2}{n^3} + \frac{A^2}{n^4}\right) \\
&\leq \mathcal{C}\frac{\mathrm{C}_\varphi^2 \mathrm{v}_{f|\varphi}^2}{Kn^2}\left(\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^4(\mathrm{w}^4)} \vee \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^\infty(\mathrm{w})}\right)\left(1 + \frac{\delta_k^2}{n} + \frac{\delta_k^4(2k)}{n^2}\right).
\end{aligned}$$

which shows the result. □

Lemma 3.7.3 (Constants for the bounded U-statistic)**:**
For $k \in \mathbb{N}$ and $\delta_k \in \mathbb{R}_+$ the real-valued, bounded and symmetric kernel $h_k^b\colon \mathbb{R}^2 \to \mathbb{R}$ given in Equation (3.36) fulfills $\mathbb{E}_g[h_k^b(Y_1, y)] = 0$ for all $y \in \mathbb{R}_+$. Under Assumption 3.1.1 the following quantities satisfy the conditions in Lemma 3.6.2:

$$\begin{aligned}
A &:= 4\delta_k^2 \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^2)}, \\
B^2 &:= 4\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\delta_k^2 \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^4(\mathrm{w}^4)}, \\
C^2 &:= \mathrm{C}_\varphi \mathrm{v}_{f|\varphi}^2 \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^4(\mathrm{w}^4)}, \\
D &:= 4\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^\infty(\mathrm{w})}.
\end{aligned}$$

Proof of Lemma 3.7.3. We compute the quantities A, B, C and D verifying the four inequalities of Lemma 3.6.2. Consider A. Since $|\psi_k^b(y,t) - \mathbb{E}_g[\psi_k^b(Y_1,t)]| \leq 2\delta_k$ for all $t \in \mathbb{R}$ and $y \in \mathbb{R}_+$ we have

$$\sup_{x,y\in\mathbb{R}} |h_k^b(x,y)\,| \leq 4\delta_k^2 \int_{-k}^{k} \frac{1}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t) = A.$$

Consider B. Using that the U-statistic is canonical and Lemma 1.2.5 (i) we get for $y \in \mathbb{R}_+$

$$\begin{aligned}
\mathbb{E}_g[|h_k^b(y,Y_1)|^2] &\leq \mathbb{E}_g\left[\left|\int_{-k}^{k} \psi_k^b(Y_1,-t)\frac{(\psi_k^b(y,t) - \mathbb{E}_g[\psi_k^b(Y_1,t)])}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right] \\
&= \mathbb{E}_g\left[\mathbb{1}_{[0,\delta_k]}(Y_1^{c-1})\left|\int_{-k}^{k} Y_1^{c-1-2\pi it}\frac{(\psi_k^b(y,t) - \mathbb{E}_g[\psi_k^b(Y_1,t)])}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right] \\
&\leq \mathbb{E}_g\left[\left|\int_{-k}^{k} Y_1^{c-1-2\pi it}\frac{(\psi_k^b(y,t) - \mathbb{E}_g[\psi_k^b(Y_1,t)])}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right] \\
&\leq \|g\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})}\int_{-k}^{k} \frac{|\psi_k^b(y,t) - \mathbb{E}_g[\psi_k^b(Y_1,t)]|^2}{|\mathcal{M}_c[\varphi](t)|^4}\,\mathrm{w}^4(t)d\lambda(t). \qquad (3.44)
\end{aligned}$$

Consequently, $|\psi_k^b(y,t) - \mathbb{E}_g[\psi_k^b(Y_1,t)]| \leq 2\delta_k$ for all $t \in \mathbb{R}$ and $y \in \mathbb{R}_+$ and Equation (3.9) implies

$$\sup_{y\in\mathbb{R}_+} \mathbb{E}_g[|h_k^b(y,Y_1)|^2] \leq \mathrm{C}_\varphi \mathrm{v}_{f|\varphi} 4\delta_k^2 \int_{-k}^{k} \frac{1}{|\mathcal{M}_c[\varphi](t)|^4}\,\mathrm{w}^4(t)d\lambda(t) = B^2.$$

Consider C. Keep in mind that for all $t \in \mathbb{R}$

$$\mathbb{E}_g\left[|\psi_k^b(Y_1,t) - \mathbb{E}_g[\psi_k^b(Y_1,t)]|^2\right] \leq \mathbb{E}_g\left[|\psi_k^b(Y_1,t)|^2\right]$$
$$= \mathbb{E}_g\left[\mathbb{1}_{[0,\delta_k]}(Y_1^{c-1})|Y_1^{c-1+2\pi i t}|^2\right] \leq \mathbb{E}_g[Y_1^{2c-2}] \leq \mathrm{v}_{f|\varphi}.$$

Using additionally the calculations of Equation (3.44) for B and Equation (3.9) again, it follows

$$\mathbb{E}_g^2\left[|h_k^b(Y_1,Y_2)|^2\right] \leq \|g\|_{\mathbb{L}_+^\infty(\mathrm{x}^{2c-1})}\int_{-k}^k \mathbb{E}_g[|\psi_k^b(y,t) - \mathbb{E}_g[\psi_k^b(Y_1,t)]|^2]\frac{\mathrm{w}^4(t)}{|\mathcal{M}_c[\varphi](t)|^4}d\lambda(t)$$
$$\leq \mathrm{C}_\varphi \mathrm{v}_{f|\varphi}^2\int_{-k}^k \frac{1}{|\mathcal{M}_c[\varphi](t)|^4}\,\mathrm{w}^4(t)d\lambda(t) = C^2.$$

Finally, consider D. $D = C$ satisfies the condition, but we find a sharper bound. For this, we first see that for all $\xi \in \mathbb{L}_+^2(g)$ due to Lemma 1.2.5 (ii) we have

$$\int_{\mathbb{R}}\left|\mathbb{E}_g\left[\xi(Y_1)\psi_k^b(Y_1,t)\right]\right|^2 d\lambda(t) = \int_{\mathbb{R}}\left|\mathbb{E}_g\left[Y_1^{c-1+2\pi i t}\xi(Y_1)\mathbb{1}_{[0,\delta_k]}(Y_1^{c-1})\right]\right|^2 d\lambda(t)$$
$$\leq \|g\|_{\mathbb{L}_+^\infty(\mathrm{x}^{2c-1})}\mathbb{E}_g\left[|\xi(Y_1)|^2\mathbb{1}_{[0,\delta_k]}(Y_1^{c-1})\right]$$
$$\leq \mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\mathbb{E}_g[|\xi(Y_1)|^2]$$

where we used again Equation (3.9). Similarly, by Lemma 1.2.5 (ii) we get

$$\int_{\mathbb{R}}|\mathbb{E}_g[Y_1^{c-1+2\pi i t}\mathbb{1}_{[0,\delta_k]}(Y_1^{c-1})]|^2 d\lambda(t) \leq \mathrm{C}_\varphi \mathrm{v}_{f|\varphi},$$

which in turn implies

$$\int_{\mathbb{R}}\left|\mathbb{E}_g\left[\xi(Y_1)\mathbb{E}_g[\psi_k^b(Y_1,t)]\right]\right|^2 d\lambda(t) \leq \int_{\mathbb{R}}\mathbb{E}_g\left[|\xi(Y_1)|^2\right]|\mathbb{E}_g[\psi_k^b(Y_1,t)]|^2 d\lambda(t)$$
$$= \mathbb{E}_g\left[|\xi(Y_1)|^2\right]\int_{\mathbb{R}}|\mathbb{E}_g[Y_1^{c-1+2\pi i t}\mathbb{1}_{[0,\delta_k]}(Y_1^{c-1})]|^2 d\lambda(t) \leq \mathbb{E}_g\left[|\xi(Y_1)|^2\right]\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}.$$

Combining the last bounds it follows

$$\int_{\mathbb{R}}|\mathbb{E}_g[\xi(Y_1)(\psi_k^b(Y_1,t) - \mathbb{E}_g[\psi_k^b(Y_1,t)])]|^2 d\lambda(t) \leq 4\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\mathbb{E}_g[|\xi(Y_1)|^2].$$

Consequently, for all $\xi, \zeta \in \mathbb{L}_+^2(g)$ with $\mathbb{E}_g[|\xi(Y_1)|^2] \leq 1$ and $\mathbb{E}_g[|\zeta(Y_1)|^2] \leq 1$ it follows

$$\int_{\mathbb{R}_+} h_k^b(x,y)\xi(x)\zeta(y)\mathbb{P}_g(dx)\mathbb{P}_g(dy)$$
$$= \int_{-k}^k \frac{\mathbb{E}_g[\xi(Y_1)(\psi_k^b(Y_1,t) - \mathbb{E}_g[\psi_k^b(Y_1,t)])]\mathbb{E}_g[\zeta(Y_2)(\psi_k^b(Y_2,-t) - \mathbb{E}_g[\psi_k^b(Y_1,-t)])]}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t)$$
$$\leq \sup_{\substack{\xi\in\mathbb{L}_+^2(g),\\ \mathbb{E}_g[|\xi(Y_1)|^2]\leq 1}}\left\{\int_{-k}^k \frac{|\mathbb{E}_g[\xi(Y_1)(\psi_k^b(Y_1,t) - \mathbb{E}_g[\psi_k^b(Y_1,t)])]|^2}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t)\right\}$$
$$\leq \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^\infty(\mathrm{w})}^2 \sup_{\substack{\xi\in\mathbb{L}_+^2(g),\\ \mathbb{E}_g[|\xi(Y_1)|^2]\leq 1}}\left\{\int_{-k}^k |\mathbb{E}_g[\xi(Y_1)(\psi_k^b(Y_1,t) - \mathbb{E}_g[\psi_k^b(Y_1,t)])]|^2 d\lambda(t)\right\}$$
$$\leq 4\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^\infty(\mathrm{w})}^2 = D.$$

This concludes the proof. □

Lemma 3.7.4 (Variance of the unbounded U-statistic)**:**
For $k \in \mathbb{N}$ and $\delta_k \in \mathbb{R}_+$ the real-valued, bounded and symmetric kernels $h_k^u, h_k^{bu} \colon \mathbb{R}_+^2 \to \mathbb{R}$ given in Equation (3.37) and Equation (4.41), respectively, fulfill for all $y \in \mathbb{R}_+$ that $\mathbb{E}_g[h_k^u(Y_1, y)] = 0$ and $\mathbb{E}_g[h_k^{bu}(Y_1, y)] = 0$. Under Assumption 3.1.1 the canonical U-statistics

$$U_k^u := \frac{1}{n(n-1)} \sum_{\substack{j \neq l \\ j,l \in [\![n]\!]}} h_k^u(Y_j, Y_l), \quad \text{and} \quad U_k^{bu} := \frac{1}{n(n-1)} \sum_{\substack{j \neq l \\ j,l \in [\![n]\!]}} h_k^{bu}(Y_j, Y_l),$$

satisfy that

$$\mathbb{E}_g[|U_k^u|^2] \leq \frac{2\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}}{n^2} \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^4(\mathrm{w}^4)}^4 \delta_k^{-2} \mathbb{E}_g[Y_1^{4(c-1)}],$$
$$\mathbb{E}_g[|U_k^{bu}|^2] \leq \frac{8\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}}{n^2} \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^4(\mathrm{w}^4)}^4 \delta_k^{-2} \mathbb{E}_g[Y_1^{4(c-1)}].$$

Proof of Lemma 3.7.4. Applying Lemma 3.6.1 we get $\mathbb{E}_g^n[|U_k^u|^2] \leq \frac{2}{n^2}\mathbb{E}_g^2[|h_k^u(Y_1, Y_2)|^2]$. With Lemma 1.2.5 (i) we have analogously to the calculations in Equation (3.44) in Lemma 3.7.3 that

$$\mathbb{E}_g[|h_k^u(y, Y_1)|^2] \leq \|g\|_{\mathbb{L}_+^\infty(\mathrm{x}^{2c-1})} \int_{-k}^{k} \frac{|\psi_k^u(y,t) - \mathbb{E}_g[\psi_k^u(Y_1,t)]|^2}{|\mathcal{M}_c[\varphi](t)|^4} \mathrm{w}^4(t) d\lambda(t).$$

Following the computations of the proof of Lemma 3.7.3, for all $t \in \mathbb{R}$ we have that

$$\begin{aligned}\mathbb{E}_g\left[|\psi_k^u(Y_1,t) - \mathbb{E}_g[\psi_k^u(Y_1,t)]|^2\right] &\leq \mathbb{E}_g\left[|\psi_k^u(Y_1,t)|^2\right] \\ &= \mathbb{E}_g\left[\mathbb{1}_{(\delta_k,\infty)}(Y_1^{c-1}) Y_1^{2(c-1)}\right] \leq \delta_k^{-2}\mathbb{E}_g[Y_1^{4(c-1)}].\end{aligned}$$

Combining the last bounds and using Equation (3.9) yield

$$\begin{aligned}\frac{n^2}{2}\mathbb{E}_g^n[|U_k^u|^2] &\leq \mathbb{E}_g^2[|h_k^u(Y_1, Y_2)|^2] \\ &\leq \|g\|_{\mathbb{L}_+^\infty(\mathrm{x}^{2c-1})} \int_{-k}^{k} \mathbb{E}_g[|\psi_k^u(Y_1,t) - \mathbb{E}_g[\psi_k^u(Y_1,t)]|^2] \frac{\mathrm{w}^4(t)}{|\mathcal{M}_c[\varphi](t)|^4} d\lambda(t) \\ &\leq \mathrm{C}_\varphi \mathrm{v}_{f|\varphi} \delta_k^{-2} \mathbb{E}_g[Y_1^{4(c-1)}] \int_{-k}^{k} \frac{1}{|\mathcal{M}_c[\varphi](t)|^4} \mathrm{w}^4(t) d\lambda(t). \end{aligned} \tag{3.45}$$

Rearranging the last inequality implies the first result. Applying Lemma 3.6.1 we get that $\mathbb{E}_g^n[|U_k^{bu}|^2] \leq \frac{2}{n^2}\mathbb{E}_g^2[|h_k^{bu}(Y_1, Y_2)|^2]$. Now, set

$$\tilde{h}_k^{ub}(x,y) := \int_{-k}^{k} \frac{(\psi_k^u(y,t) - \mathbb{E}_g[\psi_k^u(Y_1,t)])(\psi_k^b(x,-t) - \mathbb{E}_g[\psi_k^b(Y_1,-t)])}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t)$$

and $\tilde{h}_k^{bu}(x,y)$ similarly. Then, we have $h_k^{bu}(x,y) := \tilde{h}_k^{bu}(x,y) + \tilde{h}_k^{ub}(x,y)$ for any $x, y \in \mathbb{R}_+$. It follows that $\mathbb{E}_g^n[|U_k^{bu}|^2] \leq \frac{4}{n^2}\mathbb{E}_g^2[|\tilde{h}_k^{bu}(Y_1, Y_2)|^2] + \frac{4}{n^2}\mathbb{E}_g^2[|\tilde{h}_k^{ub}(Y_1, Y_2)|^2]$. Similarly to Equation (3.44) in Lemma 3.7.3 we get

$$\begin{aligned}&\mathbb{E}_g[|\tilde{h}_k^{bu}(Y_1, y)|^2] + \mathbb{E}_g[|\tilde{h}_k^{ub}(y, Y_1)|^2] \\ &\leq 2\|g\|_{\mathbb{L}_+^\infty(\mathrm{x}^{2c-1})} \int_{-k}^{k} \frac{|\psi_k^u(y,t) - \mathbb{E}_g[\psi_k^u(Y_1,t)]|^2}{|\mathcal{M}_c[\varphi](t)|^4} \mathrm{w}^4(t) d\lambda(t).\end{aligned}$$

Analogously to Equation (3.45) it follows

$$\begin{aligned}\frac{n^2}{4}\mathbb{E}_g^n[|U_k^{bu}|^2] &\leq \mathbb{E}_g[|\tilde{h}_k^{bu}(Y_1,y)|^2] + \mathbb{E}_g[|\tilde{h}_k^{ub}(y,Y_1)|^2] \\ &\leq 2\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\delta_k^{-2}\mathbb{E}_g[Y_1^{4(c-1)}]\int_{-k}^{k}\frac{1}{|\mathcal{M}_c[\varphi](t)|^4}\,\mathrm{w}^4(t)d\lambda(t).\end{aligned}$$

Rearranging the last inequality implies the second result. □

3.7.2 Linear statistic results

Lemma 3.7.5 (Concentration of the linear statistic)**:**
Under Assumption 3.1.1 there exists a universal numerical constant $\mathcal{C}$ such that

$$\begin{aligned}&\mathbb{E}_g^n\left[\max_{k<k'\leq M}\left(|W_{k'}-W_k|^2-(\frac{1}{32}V(k')+\frac{1}{16}(q_k^2-q^2)^2)\right)_+\right] \\ &\leq \mathcal{C}(\mathrm{C}_\varphi^2\mathrm{v}_{f|\varphi}^2+(\mathbb{E}_g[Y_1^{4(c-1)}])^2)\frac{1}{n}+\mathcal{C}(q_k^2-q^2)^2.\end{aligned}$$

for W_k defined in Equation (3.12), $V(k)$ in Equation (3.26) and $\kappa \geq 64^2 * 4 * 128^2$.

Proof of Lemma 3.7.5. For $k \in \mathbb{N}$ consider W_k defined in Equation (3.12). We intend to apply the concentration inequality Lemma 3.6.3 where we need to compute the quantities b and v verifying the required inequalities. Analogously to the U-statistic part, note that $W_{k'} - W_k$ is generally not bounded since $|x^{c-1+it}|$ is not bounded for $x \in \mathbb{R}_+$. Therefore, we use again the notation ψ_k^b and ψ_k^u analogously to Equation (3.35), i.e., given a sequence $(\delta_k)_{k\in\mathbb{N}}$ define for $k \in \mathbb{N}$ for $y \in \mathbb{R}_+$ and $t \in \mathbb{R}$

$$\psi_k^b(y,t) := \mathbb{1}_{[0,\delta_k]}(y^{c-1})y^{c-1+2\pi it} \quad\text{and}\quad \psi_k^u(y,t) := \mathbb{1}_{(\delta_k,\infty)}(y^{c-1})y^{c-1+2\pi it}.$$

Then, for $k,k' \in [\![M]\!]$ with $k \leq k'$ we define

$$W_{k',k}^b := \frac{1}{n}\sum_{j\in[\![n]\!]}\int_{[-k',k']\setminus[-k,k]}(\psi_{k'}^b(Y_j,t)-\mathbb{E}_g[\psi_{k'}^b(Y_1,t)])\frac{\mathcal{M}_c[g](-t)}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t), \tag{3.46}$$

$$W_{k',k}^u := \frac{1}{n}\sum_{j\in[\![n]\!]}\int_{[-k',k']\setminus[-k,k]}(\psi_{k'}^u(Y_j,t)-\mathbb{E}_g[\psi_{k'}^u(Y_1,t)])\frac{\mathcal{M}_c[g](-t)}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t). \tag{3.47}$$

We evidently have $Y_j^{c-1+2\pi it} = \psi_{k'}^b(Y_j,t) + \psi_{k'}^u(Y_j,t)$ for any $k' \in [\![M]\!]$ and, consequently, $W_{k'} - W_k = W_{k',k}^b + W_{k',k}^u$ for any $k \in [\![M]\!]$ with $k \leq k'$. Note that we choose the cut-off of the random variable Y_j only in dependence of the larger dimension parameter $k' \in [\![M]\!]$. With this, we immediately obtain

$$\begin{aligned}&\mathbb{E}_g^n\left[\max_{k<k'\leq M}\left(|W_{k'}-W_k|^2-(\frac{1}{32}V(k')+\frac{1}{16}(q_k^2-q^2)^2)\right)_+\right] \\ &\leq 2\sum_{k<k'\leq M}\mathbb{E}_g^n\left[\left(|W_{k',k}^b|^2-(\frac{1}{64}V(k')+\frac{1}{32}(q_k^2-q^2)^2)\right)_+\right] \\ &\quad + 2\sum_{k<k'\leq M}\mathbb{E}_g^n[|W_{k',k}^u|^2].\end{aligned} \tag{3.48}$$

We begin by considering the first summand. In Lemma 3.7.6 we apply the Bernstein inequality Lemma 3.6.3 to $W^b_{k',k}$. To be more precise, with Ω_k defined in Equation (3.27) and constants $\kappa^*, d > 0$ as in Lemma 3.6.2 we introduce

$$\delta^2_{k'} := \Omega_{k'} \quad \text{and} \quad K := \Omega^2_{k'}. \tag{3.49}$$

Further, set

$$V^b_2(k') := (\log K)\|(\mathbb{1}_{[-k',k']} - \mathbb{1}_{[-k,k]})\mathcal{M}_c[f]/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^4)} \cdot \frac{1}{n}\left(4\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} + \frac{128\delta^2_{k'}(2k')(\log K)}{n}\right).$$

We show below in Equation (3.53) that $V^b_2(k) \leq \frac{1}{64}V(k) + \frac{1}{32}(q^2_k - q^2)^2$. With this we get from Lemma 3.7.6 that

$$\begin{aligned}
&\mathbb{E}^n_g\left[\left(|W^b_{k',k}|^2 - (\frac{1}{64}V(k') + \frac{1}{32}(q^2_k - q^2)^2)\right)_+\right] \leq \mathbb{E}^n_g\left[\left(|W^b_{k',k}|^2 - V^b_2(k')\right)_+\right] \\
&\leq \frac{\mathcal{C}}{K}\|(\mathbb{1}_{[-k',k']} - \mathbb{1}_{[-k,k]})\mathcal{M}_c[f]/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^4)}\frac{1}{n}\left(\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} + \frac{\delta^2_{k'}(2k')}{n}\right).
\end{aligned}$$

For $a, b \in \mathbb{R}$ it holds $2ab \leq a^2 + b^2$

$$\begin{aligned}
&2\frac{1}{K}\|(\mathbb{1}_{[-k',k']} - \mathbb{1}_{[-k,k]})\mathcal{M}_c[f]/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^4)}\frac{1}{n}\left(\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} + \frac{\delta^2_{k'}(2k')}{n}\right) \\
&\leq 2\frac{1}{K}\|\mathbb{1}_{[-k',k']}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^\infty(\mathrm{w})}\|\mathbb{1}_{\mathbb{R}\setminus[-k,k]}\mathcal{M}_c[f]\|^2_{\mathbb{L}^2(\mathrm{w}^2)}\frac{1}{n}\left(\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} + \frac{\delta^2_{k'}(2k')}{n}\right) \\
&\leq \frac{1}{K^2}k'^2\|\mathbb{1}_{[-k',k']}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^\infty(\mathrm{w})}\frac{1}{n^2}\left(\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} + \frac{\delta^2_{k'}(2k')}{n}\right)^2 \\
&\qquad + \frac{1}{k'^2}\|\mathbb{1}_{\mathbb{R}\setminus[-k,k]}\mathcal{M}_c[f]\|^4_{\mathbb{L}^2(\mathrm{w}^2)} \\
&\leq \mathcal{C}\mathrm{C}^2_\varphi \mathrm{v}^2_{f|\varphi}\frac{1}{k'^2 n}\frac{1}{K^2}\frac{1}{n}\|\mathbb{1}_{[-k',k']}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^\infty(\mathrm{w})}(2k')^6\delta^4_{k'} + \frac{1}{k'^2}(q^2_k - q^2)^2.
\end{aligned} \tag{3.50}$$

Since by definition of Ω_k in Equation (3.27) and by Equation (3.49) we have that

$$\frac{1}{n}\|\mathbb{1}_{[-k',k']}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^\infty(\mathrm{w})}(2k')^6 \leq \Omega^2_{k'} = \frac{K^2}{\delta^4_{k'}}. \tag{3.51}$$

Combining the last inequalities and since $\sum_{k'\in\mathbb{N}} k'^{-2} = \pi^2/6$ we get

$$\begin{aligned}
&\sum_{k<k'\leq M}\mathbb{E}^n_g\left[\left(|W^b_{k',k}|^2 - (\frac{1}{64}V(k') + \frac{1}{32}(q^2_k - q^2)^2)\right)_+\right] \\
&\leq \mathcal{C}\mathrm{C}^2_\varphi \mathrm{v}^2_{f|\varphi}\frac{1}{n}\sum_{k'\in\mathbb{N}}\frac{1}{k'^2} + (q^2_k - q^2)^2\sum_{k'\in\mathbb{N}}\frac{1}{k'^2} \\
&\leq \mathcal{C}\mathrm{C}^2_\varphi \mathrm{v}^2_{f|\varphi}\frac{1}{n} + \mathcal{C}(q^2_k - q^2)^2.
\end{aligned} \tag{3.52}$$

To conclude the calculation for the first summand of Equation (3.48), we show now that it holds $V_2^b(k) \leq \frac{1}{64}V(k) + \frac{1}{32}(q_k^2 - q^2)^2$. For this, with similar calculations as in Equation (3.50) we see with Equation (3.49) that

$$
\begin{aligned}
&V_2^b(k') \\
&= (\log K)\|(\mathbb{1}_{[-k',k']} - \mathbb{1}_{[-k,k]})\mathcal{M}_c[f]/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^4)} \\
&\quad \cdot \frac{1}{n}\left(4\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} + \frac{128\delta_{k'}^2(2k')(\log K)}{n}\right) \\
&\leq \frac{2}{n}\log(\Omega_{k'})\|\mathbb{1}_{[-k',k']}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^\infty(\mathrm{w})}\|(\mathbb{1}_{[-k',k']} - \mathbb{1}_{[-k,k]})\mathcal{M}_c[f]\|^2_{\mathbb{L}^2(\mathrm{w}^2)} \\
&\quad \cdot \left(4\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} + \frac{128\Omega_{k'}(2k')2(\log \Omega_{k'})}{n}\right) \\
&\leq 64\frac{(4\mathrm{C}_\varphi \mathrm{v}_{f|\varphi})^2}{n^2}(\log \Omega_{k'})^2\|\mathbb{1}_{[-k',k']}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^\infty(\mathrm{w})} \\
&\quad + \frac{1}{64}\|(\mathbb{1}_{[-k',k']} - \mathbb{1}_{[-k,k]})\mathcal{M}_c[f]\|^4_{\mathbb{L}^2(\mathrm{w}^2)} \\
&\quad + 64\frac{1}{n^2}(\log \Omega_{k'})^2\|\mathbb{1}_{[-k',k']}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^\infty(\mathrm{w})}\frac{4 * 128^2\Omega_{k'}^2(2k')^2(\log \Omega_{k'})^2}{n^2} \\
&\quad + \frac{1}{64}\|(\mathbb{1}_{[-k',k']} - \mathbb{1}_{[-k,k]})\mathcal{M}_c[f]\|^4_{\mathbb{L}^2(\mathrm{w}^2)} \\
&\leq 64\frac{(\log \Omega_{k'})^2}{n^2}\|\mathbb{1}_{[-k',k']}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^\infty(\mathrm{w})}\left((4\mathrm{C}_\varphi \mathrm{v}_{f|\varphi})^2 + \frac{4 * 128^2\Omega_{k'}^2(2k')^2(\log \Omega_{k'})^2}{n^2}\right) \\
&\quad + \frac{1}{32}(q_k^2 - q^2)^2.
\end{aligned}
$$

Exploiting the definition of $V(k')$, see Equation (3.26), it follows

$$
V_2^b(k') \leq \frac{64 * 4 * 128^2}{\kappa}V(k') + \frac{1}{32}(q_k^2 - q^2)^2 \leq \frac{1}{64}V(k') + \frac{1}{32}(q_k^2 - q^2)^2 \tag{3.53}
$$

since $\kappa \geq 64^2 * 4 * 128^2$. For the unbounded part, i.e., the second summand of Equation (3.48), we get with Lemma 3.7.7, with similar calculations as in Equation (3.50), Equation (3.52) and using Equation (3.51) that

$$
\begin{aligned}
&\sum_{k<k'\leq M} \mathbb{E}_g[|W^u_{k',k}|^2] \\
&\leq \sum_{k<k'\leq M} \frac{1}{n}\|(\mathbb{1}_{[-k',k']} - \mathbb{1}_{[-k,k]})\mathcal{M}_c[f]/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^4)}(2k')\delta_{k'}^{-2}\mathbb{E}_g[Y_1^{4(c-1)}] \\
&\leq \sum_{k'\in\mathbb{N}}\left((\mathbb{E}_g[Y_1^{4(c-1)}])^2\frac{k'^4}{\Omega_{k'}^2 n^2}\|\mathbb{1}_{[-k',k']}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^\infty(\mathrm{w})} + \frac{1}{k'^2}\|\mathbb{1}_{\mathbb{R}\setminus[-k,k]}\mathcal{M}_c[f]\|^4_{\mathbb{L}^2(\mathrm{w}^2)}\right) \\
&\leq \mathcal{C}(\mathbb{E}_g[Y_1^{4(c-1)}])^2\frac{1}{n} + \mathcal{C}(q_k^2 - q^2)^2.
\end{aligned}
$$

Inserting now the last bound and Equation (3.52) into Equation (3.48) yields the result. □

Lemma 3.7.6 (Concentration of the bounded linear statistic)**:**
Let the centered linear statistic $W^b_{k',k}$ for $k, k' \in \mathbb{N}$, $k' \geq k$ and $\delta_{k'} \in \mathbb{R}_+$, be defined as in Equation (3.46). Under Assumption 3.1.1 there exists a universal numerical constant $\mathcal{C}$ such that for each $K \geq 1$

$$\mathbb{E}^n_g\left[\left(|W^b_{k',k}|^2 - V^b_2(k')\right)_+\right] \leq \frac{\mathcal{C}}{K}\|(\mathbb{1}_{[-k',k']} - \mathbb{1}_{[-k,k]})\mathcal{M}_c[f]/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^4)}\frac{1}{n}\left(\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} + \frac{\delta^2_{k'}(2k')}{n}\right)$$

with C_φ, $\mathrm{v}_{f|\varphi}$ defined in Equation (3.8) and

$$V^b_2(k') := (\log K)\|(\mathbb{1}_{[-k',k']} - \mathbb{1}_{[-k,k]})\mathcal{M}_c[f]/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^4)} \cdot \frac{1}{n}\left(4\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} + \frac{128\delta^2_{k'}(2k')(\log K)}{n}\right).$$

Proof of Lemma 3.7.6. In the following, for $k \leq k'$ denote $\mathrm{I}_{k',k} := \mathbb{1}_{[-k',k']} - \mathbb{1}_{[-k,k]}$. First, for $i \in [\![n]\!]$ define the i.i.d. random variables

$$Z_i := \int_{\mathbb{R}} (\psi^b_{k'}(Y_i,t) - \mathbb{E}_g[\psi^b_{k'}(Y_1,t)])\frac{\mathrm{I}_{k',k}(t)\mathcal{M}_c[f](-t)}{\mathcal{M}_c[\varphi](t)}\,\mathrm{w}^2(t)d\lambda(t)$$

where $\mathbb{E}_g[Z_i] = 0$ and $W^b_{k',k} = \frac{1}{n}\sum_{i\in[\![n]\!]} Z_i$. We intend to apply Lemma 3.6.3 which needs quantities $\mathbb{E}_g[|Z_i|^2] \leq v$ and $|Z_i| \leq b$. Consider b first. Since $|\psi^b_{k'}(y,t) - \mathbb{E}_g[\psi^b_{k'}(Y_1,t)]| \leq 2\delta_{k'}$ for all $t \in \mathbb{R}$ and $y \in \mathbb{R}_+$ we have

$$\begin{aligned}|Z_i|^2 &\leq 4\delta^2_{k'}\left|\int_{\mathbb{R}} \mathrm{I}_{k',k}(t)\frac{|\mathcal{M}_c[f](t)|}{|\mathcal{M}_c[\varphi](t)|}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\\ &\leq 4\delta^2_{k'}(2k')\int_{\mathbb{R}} |\,\mathrm{I}_{k',k}(t)|^2\left|\frac{\mathcal{M}_c[f](t)}{\mathcal{M}_c[\varphi](t)}\right|^2 \mathrm{w}^4(t)d\lambda(t)\\ &= 4\delta^2_{k'}(2k')\|\,\mathrm{I}_{k',k}\,\mathcal{M}_c[f]/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^4)} := b^2.\end{aligned}$$

Secondly, consider v. Using Lemma 1.2.5 (i) we get that

$$\begin{aligned}\mathbb{E}_g\left[|Z_i|^2\right] &\leq \mathbb{E}_g\left[\left|\int_{\mathbb{R}} \psi^b_{k'}(Y_i,t)\frac{\mathrm{I}_{k',k}(t)\mathcal{M}_c[f](-t)}{\mathcal{M}_c[\varphi](t)}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right]\\ &\leq \mathbb{E}_g\left[\left|\int_{\mathbb{R}} Y_i^{c-1+2\pi i t}\frac{\mathrm{I}_{k',k}(t)\mathcal{M}_c[f](-t)}{\mathcal{M}_c[\varphi](t)}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right]\\ &\leq \|g\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})}\int_{\mathbb{R}}\frac{\mathrm{I}_{k',k}(t)|\mathcal{M}_c[f](t)|^2}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^4(t)d\lambda(t).\end{aligned}$$

Using Equation (3.9) we get $\mathbb{E}_g\left[|Z_i|^2\right] \leq \mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\|\,\mathrm{I}_{k',k}\,\mathcal{M}_c[f]/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^4)} =: v$. Consequently, we have that

$$\begin{aligned}&n^{-1}(4v + 32b^2(\log K)n^{-1})(\log K)\\ &= n^{-1}(\log K)\|\,\mathrm{I}_{k',k}\,\mathcal{M}_c[f]/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^4)}\left(4\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} + 128\delta^2_{k'}(2k')(\log K)n^{-1}\right)\\ &= V^b_2(k').\end{aligned}$$

Consequently by Lemma 3.6.3 there exists a universal numerical constant $\mathcal{C}$ such that

$$
\begin{aligned}
&n\mathbb{E}_g^n\left[\left(|W_{k',k}^b|^2 - V_2^b(k')\right)_+\right]\\
&= \mathbb{E}_g^n\left[\left(\left|\sqrt{n}W_{k',k}^b\right|^2 - (4v + 32b^2(\log K)n^{-1})(\log K)\right)_+\right]\\
&\leq \frac{8}{K}(v + 16b^2n^{-1}) = \frac{\mathcal{C}}{K}\| \mathrm{I}_{k',k}\, \mathcal{M}_c[f]/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^4)} \left(\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} + (2k')\delta_{k'}^2 n^{-1}\right).
\end{aligned}
$$

This shows the claim and concludes the proof. □

Lemma 3.7.7 (Variance of the unbounded linear statistic)**:**
Consider the centered linear statistic $W_{k',k}^u$ for $k, k' \in \mathbb{N}$, $k' \geq k$ and $\delta_{k'} \in \mathbb{R}_+$, defined as in Equation (3.47). Under Assumption 3.1.1 we have that

$$
\mathbb{E}_g^n\left[|W_{k',k}^u|^2\right] \leq \frac{1}{n}\|(\mathbb{1}_{[-k',k']} - \mathbb{1}_{[-k,k]})\mathcal{M}_c[f]/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^4)}(2k')\delta_{k'}^{-2}\mathbb{E}_g[Y_1^{4(c-1)}].
$$

Proof of Lemma 3.7.7. We use again the notation $\mathrm{I}_{k',k} := \mathbb{1}_{[-k',k']} - \mathbb{1}_{[-k,k]}$ for $k \leq k'$. By independence of $(Y_j)_{j\in[\![n]\!]}$ and applying Cauchy–Schwarz inequality we have that

$$
\begin{aligned}
n\mathbb{E}_g^n[|W_{k',k}^u|^2] &= \mathbb{E}_g^n\left[\left|\int_{\mathbb{R}}(\psi_{k'}^u(Y_1,t) - \mathbb{E}_g[\psi_{k'}^u(Y_1,t)])\frac{\mathrm{I}_{k',k}(t)\mathcal{M}_c[f](-t)}{\mathcal{M}_c[\varphi](t)}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right]\\
&\leq \mathbb{E}_g^n\left[\left|\int_{\mathbb{R}}\mathrm{I}_{k',k}(t)\psi_{k'}^u(Y_1,t)\frac{\mathrm{I}_{k',k}(t)\mathcal{M}_c[f](-t)}{\mathcal{M}_c[\varphi](t)}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right]\\
&\leq \left(\int_{\mathbb{R}}\mathrm{I}_{k',k}(t)\mathbb{E}_g[|\psi_{k'}^u(Y_1,t)|^2]d\lambda(t)\right)\left(\int_{\mathbb{R}}\frac{\mathrm{I}_{k',k}(t)|\mathcal{M}_c[f](t)|^2}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^4(t)d\lambda(t)\right).
\end{aligned}
$$

Moreover, for all $t \in \mathbb{R}$ it holds

$$
\mathbb{E}_g[|\psi_{k'}^u(Y_1,t)|^2] = \mathbb{E}_g[\mathbb{1}_{(\delta_{k'},\infty)}(Y_1^{c-1})Y_1^{2(c-1)}] \leq \delta_{k'}^{-2}\mathbb{E}_g[Y_1^{4(c-1)}].
$$

Since $\int_{\mathbb{R}}\mathrm{I}_{k',k}(t)d\lambda(t) \leq 2k'$ the last bound implies the result. □

CHAPTER 4

Hypothesis testing under multiplicative measurement errors

In this chapter, we consider nonparametric adaptive goodness-of-fit hypothesis testing for the density of a strictly positive random variable X in a multiplicative measurement error model. More precisely, we have access to an independent and identically distributed (i.i.d.) sample of size $n \in \mathbb{N}$ from $Y = XU$, where X and U are strictly positive independent random variables. We denote the unknown density of X by f and assume that U admits a known density φ with respect to the Lebesgue measure λ_+ on the positive real line $\mathbb{R}_+$. We refer to the density of Y by g, which is then given by the multiplicative convolution of f and φ, i.e., for $y \in \mathbb{R}_+$

$$g(y) = (f \circledast \varphi)(y) := \int_{\mathbb{R}_+} f(x)\varphi(y/x)x^{-1}d\lambda_+(x).$$

Consequently, inference is based on observations of Y generated by a transformation $g = T(f)$ of the density f given by a multiplicative convolution with the error density φ. Throughout this chapter, whenever we use the term convolution, we have the multiplicative convolution in mind. While the last chapters dealt with nonparametric estimation, now, we consider hypothesis testing. More precisely, for a given density f_o we are interested in a nonparametric goodness-of-fit test, that is, we want to decide whether

$$H_0 \colon f = f_o \quad \text{against} \quad H_1 \colon f \neq f_o,$$

based on i.i.d. observations $(Y_j)_{j \in [\![n]\!]}$ of Y.

Related literature Concerning the testing task, in the literature there exist several definitions of rates and radii of testing in an asymptotic and nonasymptotic sense. The classical definition of an asymptotic rate of testing for nonparametric alternatives is introduced in the series of papers by Ingster (1993b,c,a). For fixed noise levels, two alternative definitions of a nonasymptotic radius of testing are typically considered. For prescribed error probabilities $\alpha, \beta \in (0,1)$, Baraud (2002), Laurent et al. (2012) and Marteau and Sapatinas (2017), amongst others, define a nonasymptotic radius of testing as the smallest separation radius ρ such that there is an α-test with maximal type II error probability over the ρ-separated alternative smaller than β. The definition we use in this chapter, which is based on the sum of both error probabilities, is adapted e.g. from Collier et al. (2017).

It should also be mentioned that the test problem yields a lower bound for the quadratic functional estimation problem. For the additive measurement error model on the real line, we refer to Butucea (2007). and Schluttenhofer and Johannes (2020b,a) extensively discuss the link between testing theory and quadratic functional estimation.

For the additive measurement error model on the real line, goodness-of-fit testing is considered, for example, in Butucea (2007). Circular observations are studied in Schluttenhofer and Johannes (2020b,a).

In this work we extend hypothesis testing to the multiplicative measurement error model. Multiplicative censoring, which corresponds to the multiplicative measurement error model with U being uniformly distributed on $[0, 1]$, has been introduced and studied by Vardi (1989) and Vardi and Zhang (1992). van Es et al. (2000) explain and motivate the use of multiplicative censoring models in survival analysis. Concerning inference in this model, there has been extensive work on estimation of the unknown density f. To name a few: Andersen and Hansen (2001) consider series expansion methods treating the model as an inverse problem, Brunel et al. (2016) use a kernel estimator for density estimation in the multiplicative censoring model, Comte and Dion (2016) consider a projection density estimator with respect to the Laguerre basis, Belomestny et al. (2016) study a Beta-distributed error. The multiplicative measurement error model covers all of these cases of multiplicative censoring. Nonparametric density estimation in the multiplicative measurement error model has been considered by Brenner Miguel et al. (2021) with known error density using a spectral cut-off regularization and by Brenner Miguel et al. (2024) with unknown error density. Brenner Miguel (2022) considers an estimation procedure using an anisotropic spectral cut-off. Brenner Miguel et al. (2023) look at the estimation of a linear functional of the unknown density and Belomestny and Goldenshluger (2020) examine point-wise density estimation in the multiplicative measurement error model. Estimating the cumulative distribution point-wise, as considered by Belomestny et al. (2024), treats also situations without absolute continuity.

Contribution We now turn to the nonparametric adaptive goodness-of-fit testing task in the multiplicative measurement error model. To distinguish between the null hypothesis and the alternative, we consider the distance between f and f_o in the weighted norm introduced in Chapter 3, i.e., we take the difference in a (possibly weighted) $\mathbb{L}^2$-norm, but also the $\mathbb{L}^2$-norm of their derivatives or the $\mathbb{L}^2$-norm of their survival function are possible. Consequently, our work extends the results of the nonparametric goodness-of-fit testing task to the multiplicative measurement error model. Following the estimation strategy of the quadratic functional in Chapter 3, which is based on the density estimation strategy in Brenner Miguel et al. (2021), we define a test statistic and derive upper bounds for its radius of testing. Secondly, we propose an adaptive testing procedure using the Bonferroni method and again derive upper bounds for the radius of testing.

Outline of this chapter We first present the testing task, introduce notation and propose the test statistic in Section 4.1. In Section 4.2 we derive bounds for the quantiles of the proposed test statistic. In Section 4.3 upper bounds for the radius of testing are derived. Next, in Section 4.4 with the max-test an adaptive testing procedure is proposed and upper bounds of the radius of testing are shown, as well. Finally, in Section 4.5, we illustrate the results in a small simulation study.

4.1 Goodness-of-fit testing task

In the multiplicative measurement error model, we assume that the random variables admit densities with respect to the Lebesgue measure on $\mathbb{R}_+$ and that the Mellin transform of the densities exists. These assumptions are summarized in the following set of functions, i.e., we set

$$\mathscr{D} := \{h \in \mathbb{L}^1_+(\mathrm{x}^{c-1}) \cap \mathbb{L}^2_+(\mathrm{x}^{2c-1}) : h \text{ Lebesgue density on } \mathbb{R}_+\}.$$

Throughout this section, the objects of the following assumption will be fixed and are assumed to be known.

Assumption 4.1.1:
Consider the multiplicative measurement error model, an arbitrary measurable symmetric density function $\mathrm{w}\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ and $c \in \mathbb{R}$. In addition, let $f_o, \varphi \in \mathscr{D}$.

Under Assumption 4.1.1 we are interested in the goodness-of-fit testing task

$$H_0\colon f = f_o \quad \text{against} \quad H_1\colon f \neq f_o, \tag{4.1}$$

where optimality is measured in a minimax sense, see Section 1.3.2 for a review on nonparametric hypothesis testing. The performance of a test is measured by how well it is able to distinguish between the null hypothesis and elements that are in some sense separated from the null. For this, first define for a measurable and symmetric density function $\mathrm{u}\colon \mathbb{R} \to \mathbb{R}$ the class

$$\mathscr{F}^{\mathrm{u}} := \{h \in \mathbb{L}^2_+(\mathrm{x}^{2c-1}) \cap \mathbb{L}^1_+(\mathrm{x}^{c-1}) : \mathcal{M}_c[h] \in \mathbb{L}^2(\mathrm{u}^2)\}.$$

Here, $\mathcal{M}_c[h]$ denotes the Mellin transform of h, see Section 1.2 for the definition and basic properties. Now, we consider the following norm to measure the separation of densities. Given $c \in \mathbb{R}$ and density function $\mathrm{w}\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ of Assumption 4.1.1, we define for $h \in \mathscr{F}^{\mathrm{w}}$ analogously to Chapter 3 the quadratic functional

$$q^2(h) := \|\mathcal{M}_c[h]\|^2_{\mathbb{L}^2(\mathrm{w}^2)} = \int_{\mathbb{R}} |\mathcal{M}_c[h](t)|^2 \, \mathrm{w}^2(t) d\lambda(t). \tag{4.2}$$

This general quadratic functional includes a large class of possible functionals, such as the $\mathbb{L}^2_+$-norm of the density itself, its derivatives or the survival function. For a more detailed discussion of these examples, see Section 1.4.

We define an *energy set* for the separation radius $\rho \in \mathbb{R}_+$ as

$$\mathscr{F}_{\geq \rho} := \left\{h \in \mathscr{F}^{\mathrm{w}} : q^2(h) \geq \rho^2\right\}.$$

Further, we assume some kind of regularity for the densities in the alternative. Analogously to the quadratic functional estimation in Chapter 3 consider the following assumption.

Assumption 4.1.2:
(i) Let $\mathrm{s}\colon \mathbb{R} \to \mathbb{R}$ be a symmetric, monotonically increasing weight function such that $\mathrm{s}(t) \to \infty$ as $|t| \to \infty$.

(ii) Assume that the symmetric function w/s is monotonically non-increasing such that $\mathrm{w}(t)/\mathrm{s}(t) = o(1)$ as $|t| \to \infty$.

Then, for $R \in \mathbb{R}_+$ define

$$\mathscr{F}^{\mathrm{s}|R} := \left\{ h \in \mathscr{F}^{\mathrm{s}} : \|\mathrm{s}\,\mathcal{M}_c[h]\|_{\mathbb{L}^2}^2 \leq R^2 \right\} \tag{4.3}$$

and write shortly

$$\mathscr{D} \cap (f_o + \mathscr{F}^{\mathrm{s}|R} \cap \mathscr{F}_{\geq\rho}) := \{h \in \mathscr{D} : h - f_o \in \mathscr{F}^{\mathrm{s}|R} \cap \mathscr{F}_{\geq\rho}\}.$$

Consequently, under Assumptions 4.1.1 and 4.1.2 the testing task can be written as

$$H_0 \colon f = f_o \quad \text{against} \quad H_1 \colon f \in \mathscr{D} \cap (f_o + \mathscr{F}^{\mathrm{s}|R} \cap \mathscr{F}_{\geq\rho}). \tag{4.4}$$

Roughly speaking, in minimax testing one searches for the smallest ρ such that (4.4) is still testable with small error probabilities. We measure the accuracy of a test $\Delta \colon \mathbb{R}_+^n \to \{0, 1\}$ by its maximal risk defined as the sum of the type I error probability and the maximal type II error probability over the ρ-separated alternative

$$\mathcal{R}\left(\Delta|\mathscr{D}, \mathscr{F}^{\mathrm{s}|R}, \rho^2\right) = \mathbb{P}^n_{\varphi, f_o}(\Delta = 1) + \sup_{f \in \mathscr{D} \cap (f_o + \mathscr{F}^{\mathrm{s}|R} \cap \mathscr{F}_{\geq\rho})} \mathbb{P}^n_{\varphi, f}(\Delta = 0).$$

Thus, given an estimator $\hat{q}^2$ of $q^2(f - f_o)$ for some suitable value $\tau \in \mathbb{R}_+$, a test is given by

$$\Delta := \mathbb{1}_{\{\hat{q}^2 \geq \tau\}}.$$

This illustrates the connection to the estimation of $q^2(f)$ discussed in Chapter 3. For brevity, we write for $x \in \mathbb{R}_+$ and $t \in \mathbb{R}$

$$\psi(x, t) := x^{c-1+2\pi i t},$$

where evidently $\mathbb{E}_g[\psi(Y, t)] = \mathcal{M}_c[g](t)$. In addition, we use for the centered term the notation $\Psi(x, t) := \psi(x, t) - \mathbb{E}_g[\psi(Y, t)]$. Before proposing an estimator of $q^2(f - f_o)$, we need the following assumption.

Assumption 4.1.3:
$\mathcal{M}_c[\varphi](t) \neq 0$ for all $t \in \mathbb{R}$ and $\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi] \in \mathbb{L}_+^\infty(\mathrm{w})$ for all $k \in \mathbb{R}_+$.

We first see that under Assumptions 4.1.1 and 4.1.3 it holds that for any $f \in \mathscr{D}$

$$\begin{aligned} q_k^2(f - f_o) &= \|\mathbb{1}_{[-k,k]}(\mathcal{M}_c[f] - \mathcal{M}_c[f_o])\|_{\mathbb{L}^2(\mathrm{w}^2)}^2 \\ &= \|\mathbb{1}_{[-k,k]}\mathcal{M}_c[f]\|_{\mathbb{L}^2(\mathrm{w}^2)}^2 - 2\langle \mathbb{1}_{[-k,k]}\mathcal{M}_c[f], \mathbb{1}_{[-k,k]}\mathcal{M}_c[f_o]\rangle_{\mathbb{L}^2(\mathrm{w}^2)} \\ &\quad + \|\mathbb{1}_{[-k,k]}\mathcal{M}_c[f_o]\|_{\mathbb{L}^2(\mathrm{w}^2)}^2 \end{aligned}$$

and with the convolution theorem (c.f. Property 1.2.2) we get

$$\begin{aligned} q_k^2(f - f_o) = \|\mathbb{1}_{[-k,k]}\mathcal{M}_c[g]/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^2(\mathrm{w}^2)}^2 - 2\langle \mathbb{1}_{[-k,k]}\mathcal{M}_c[g]/\mathcal{M}_c[\varphi], \mathbb{1}_{[-k,k]}\mathcal{M}_c[g_o]/\mathcal{M}_c[\varphi]\rangle_{\mathbb{L}^2(\mathrm{w}^2)} \\ + \|\mathbb{1}_{[-k,k]}\mathcal{M}_c[g_o]/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^2(\mathrm{w}^2)}^2. \end{aligned} \tag{4.5}$$

Since f_o and φ are known, the third term of Equation (4.5) is known as well. For the second term, we introduce the following estimator:

$$\begin{aligned}\widehat{S}_k &:= \langle \mathbb{1}_{[-k,k]}\psi(Y_j,\cdot)/\mathcal{M}_c[\varphi], \mathbb{1}_{[-k,k]}\mathcal{M}_c[g_o]/\mathcal{M}_c[\varphi]\rangle_{\mathbb{L}^2(\mathrm{w}^2)} \\ &= \frac{1}{n}\sum_{j\in[\![n]\!]}\int_{\mathbb{R}} \mathbb{1}_{[-k,k]}(t)|\mathcal{M}_c[\varphi](t)|^{-2}\psi(Y_j,t)\overline{\mathcal{M}_c[g_o](t)}\,\mathrm{w}^2(t)d\lambda(t).\end{aligned}$$

For the first term of Equation (4.5) we introduce the estimator

$$\widehat{T}_k := \frac{1}{n(n-1)}\sum_{\substack{j\neq l\\ j,l\in[\![n]\!]}}\int_{-k}^{k}\frac{\psi(Y_j,t)\psi(Y_l,-t)}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t). \tag{4.6}$$

Due to the independence of $(Y_j)_{j\in[\![n]\!]}$ we have that

$$\begin{aligned}q_k^2(f) := \mathbb{E}_g^n[\widehat{T}_k] &= \int_{-k}^{k}\frac{\mathbb{E}_g[\psi(Y_1,t)]\mathbb{E}_g[\psi(Y_2,-t)]}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t) \\ &= \int_{-k}^{k}|\mathcal{M}_c[f](t)|^2\,\mathrm{w}^2(t)d\lambda(t).\end{aligned}$$

Consequently, the test statistic defined by

$$\hat{q}_k^2 := \widehat{T}_k - 2\widehat{S}_k + q_k^2(f_o). \tag{4.7}$$

gives an unbiased estimator $\hat{q}_k^2$ of $q_k^2(f-f_o)$. Before considering upper bounds of the radius of testing, in the next section, we derive bounds for the quantiles of the proposed test statistic.

4.2 Bounds for the quantiles of the test statistic

In this section, we derive bounds for the quantiles of the test statistic $\hat{q}_k^2$ defined in Equation (4.7). To define the critical value $\tau_k(\alpha)$ for $\alpha\in(0,1)$ of $\hat{q}_k^2$, we first introduce some notation. We set $L_\alpha := 1-\log\alpha \geq 1$ and define for $k\in\mathbb{R}_+$

$$\mathbf{v}_k := \int_{-k}^{k}\frac{1}{|\mathcal{M}_c[\varphi](t)|^4}\,\mathrm{w}^4(t)d\lambda(t), \tag{4.8}$$

$$\mathbf{m}_k := \sup_{t\in[-k,k]}\frac{1}{|\mathcal{M}_c[\varphi](t)|^4}\,\mathrm{w}^4(t). \tag{4.9}$$

Due to Assumption 4.1.3 both terms are well-defined and finite. More precisely, due to the assumption $\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\in\mathbb{L}_+^\infty(\mathrm{w})$ we have that $\mathbf{m}_k$ is finite. It immediately follows that for any $p\in\mathbb{N}$ it holds $\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\in\mathbb{L}_+^{2p}(\mathrm{w}^{2p})$ for all $k\in\mathbb{R}_+$ since

$$\int_{-k}^{k}\frac{1}{|\mathcal{M}_c[\varphi](t)|^{2p}}\,\mathrm{w}^{2p}(t)d\lambda(t)\leq\mathbf{m}_k^{2p}(2k).$$

To show bounds for the quantiles of the test statistic we need additional moment assumptions.

Assumption 4.2.1:
Let $\varphi\in\mathbb{L}_+^\infty(\mathrm{x}^{2c-1})\cap\mathbb{L}_+^1(\mathrm{x}^{4(c-1)})$ and $f_o\in\mathbb{L}_+^1(\mathrm{x}^{4(c-1)})$.

It follows immediately from Assumption 4.2.1 that $\varphi \in \mathbb{L}^1_+(\mathrm{x}^{2(c-1)})$. For $x, y \in \mathbb{R}$ use the notation $x \vee y := \max\{x, y\}$ and $x \wedge y := \min\{x, y\}$. Then, we write

$$\mathrm{C}_\varphi := \|\varphi\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})} / \|\varphi\|_{\mathbb{L}^1_+(\mathrm{x}^{2(c-1)})} \vee 1, \tag{4.10}$$

which is finite under Assumptions 4.2.1 and 4.1.3. If $h \in \mathbb{L}^1_+(\mathrm{x}^{2(c-1)})$, then due to Property 1.2.3 it holds that $h \circledast \varphi \in \mathbb{L}^1_+(\mathrm{x}^{2(c-1)})$ and we denote

$$\mathrm{v}_{h|\varphi} := \|h \circledast \varphi\|_{\mathbb{L}^1_+(\mathrm{x}^{2(c-1)})} \vee 1.$$

Analogously, if $h \in \mathbb{L}^1_+(\mathrm{x}^{4(c-1)})$, then due to Property 1.2.3 it holds that $h \circledast \varphi \in \mathbb{L}^1_+(\mathrm{x}^{4(c-1)})$ and we write

$$\overline{\mathrm{v}}_{h|\varphi} := \|h \circledast \varphi\|_{\mathbb{L}^1_+(\mathrm{x}^{4(c-1)})} \vee 1. \tag{4.11}$$

Note that under Assumption 4.2.1, the constants $\mathrm{v}_{f_o|\varphi}$ and $\overline{\mathrm{v}}_{f_o|\varphi}$ are finite. Finally, the critical value is given for $\alpha \in (0, 1)$ as

$$\tau_k(\alpha) := \left(18\mathrm{C}_\varphi \overline{\mathrm{v}}_{f_o|\varphi} + 69493\frac{\sqrt{2k}}{n}\frac{L_{\alpha/2}}{\alpha}\right) L^{1/2}_{\alpha/2}\frac{\sqrt{\mathbf{v}_k}}{n} + 52\mathrm{v}_{f_o|\varphi}\mathrm{C}_\varphi L_{\alpha/2}\frac{\sqrt{\mathbf{m}_k}}{n}. \tag{4.12}$$

The key element to analyze the behavior of the test statistic $\hat{q}^2_k$ is the following decomposition. It holds

$$\hat{q}^2_k = U_k + 2W_k + q^2_k(f - f_o) \tag{4.13}$$

with the canonical U-statistic

$$U_k := \frac{1}{n(n-1)} \sum_{\substack{j_1 \neq j_2 \\ j_1, j_2 \in [\![n]\!]}} \int_{-k}^{k} \frac{\Psi(Y_{j_1}, t)\overline{\Psi}(Y_{j_2}, t)}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t) \tag{4.14}$$

and the centered linear statistic

$$W_k := \frac{1}{n} \sum_{j \in [\![n]\!]} \int_{-k}^{k} \frac{\Psi(Y_j, t)(\overline{\mathcal{M}_c[g]}(t) - \overline{\mathcal{M}_c[g_o]}(t))}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t). \tag{4.15}$$

For the two following propositions, we apply results for the U-statistic and linear statistic, which are more technical and can be found in Section 4.7. We first give a bound under the null hypothesis.

Proposition 4.2.2 (Bound for the quantiles of $\hat{q}^2_k$ under the null hypothesis)**:**
Let Assumptions 4.1.1, 4.2.1 and 4.1.3 be satisfied and let $\alpha \in (0, 1)$, $n \geq 2$ and $k \in \mathbb{N}$. Consider the estimator $\hat{q}^2_k$ and the threshold $\tau_k(\alpha)$ defined in (4.7) and (4.12), respectively. Under the null hypothesis we have that

$$\mathbb{P}^n_{\varphi, f_o}(\hat{q}^2_k \geq \tau_k(\alpha)) \leq \alpha.$$

Proof of Proposition 4.2.2. If $f = f_o$ and, hence $g = g_o$ and $\mathcal{M}_c[g] = \mathcal{M}_c[g_o]$, the decomposition (4.13) simplifies to $\hat{q}^2_k = U_k$. Due to Assumption 4.2.1, we have that $f_o, \varphi \in \mathbb{L}^1_+(\mathrm{x}^{4(c-1)})$ and the assumptions of Lemma 4.7.4 given in Section 4.7 are satisfied for $p = 2$. Moreover, with $\gamma = \alpha$ and using the notations in (4.8) and (4.9), we have that $\tau^{U_k}(\gamma) = \tau_k(\alpha)$. Thus, we immediately obtain the result from Lemma 4.7.4. □

Let us continue with the bound of the quantiles of the test statistic under the alternative.

Proposition 4.2.3 (Bound for the quantiles of $\hat{q}_k^2$ under the alternative)**:**
Let Assumptions 4.1.1, 4.2.1 and 4.1.3 be satisfied and let $\beta \in (0,1)$, $n \geq 2$ and $k \in \mathbb{N}$. Consider the estimator $\hat{q}_k^2$ defined in (4.7). Under the alternative, if $f \in \mathscr{D} \cap \mathbb{L}_+^1(\mathrm{x}^{4(c-1)})$ satisfies for an arbitrary critical value $\tau \in \mathbb{R}_+$ the separation condition

$$q_k^2(f - f_o) \geq 2\tau + 2\left(18\mathrm{C}_\varphi \overline{\mathrm{v}}_{f|\varphi} + 138986\frac{L_{\beta/4}}{\beta}\frac{\sqrt{2k}}{n}\right) L_{\beta/4}^{1/2}\frac{\sqrt{\mathbf{v}_k}}{n} + 112\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\frac{L_{\beta/4}}{\beta}\frac{\sqrt{\mathbf{m}_k}}{n}, \tag{4.16}$$

then, we have that

$$\mathbb{P}_{\varphi,f}^n(\hat{q}_k^2 < \tau) \leq \beta.$$

Proof of Proposition 4.2.2. Keeping the decomposition (4.13) in mind, we control the deviations of the U-statistic U_k and the linear statistic V_k separately, applying Lemma 4.7.4 and Lemma 4.7.5, respectively. The assumptions of Lemma 4.7.4 are satisfied with $p = 2$. Therefore, making use of $\gamma = \beta/2$ and the notations of (4.8) and (4.9) we get for the quantile $\tau^{U_k}(\beta/2)$ that

$$\tau^{U_k}(\beta/2) = \left(18\overline{\mathrm{v}}_{f|\varphi}\mathrm{C}_\varphi + 138986\frac{L_{\beta/4}}{\beta}\frac{\sqrt{2k}}{n}\right) L_{\beta/4}^{1/2}\frac{\sqrt{\mathbf{v}_k}}{n} + 52\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} L_{\beta/4}\frac{\sqrt{\mathbf{m}_k}}{n}.$$

The event $\Omega_U := \{U_k \leq -\tau^{U_k}(\beta/2)\}$ satisfies

$$\mathbb{P}_{\varphi,f}^n(\Omega_U) = \mathbb{P}_{\varphi,f}^n(-U_k \geq \tau^{U_k}(\beta/2)) \leq \frac{\beta}{2} \tag{4.17}$$

due to Lemma 4.7.4 and the usual symmetry argument. The assumptions of Lemma 4.7.5 are satisfied and with $\gamma = \beta/2$ the quantile of the linear statistic W_k is given by

$$\tau^{W_k}(\beta/2) = 2\frac{\sqrt{\mathbf{m}_k}}{n}\frac{\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}}{\beta} + \frac{1}{4}q_k^2(f - f_o).$$

Define the event $\Omega_W := \{W_k < -\tau^{W_k}(\beta/2)\}$. Due to Lemma 4.7.5 we have

$$\mathbb{P}_{f,\varphi}^n(\Omega_W) = \mathbb{P}_{f,\varphi}^n(W_k < -\tau^{W_k}(\beta/2)) \leq \frac{\beta}{2}. \tag{4.18}$$

Further, using $L_{\beta/4}, \frac{1}{\beta} \geq 1$ it holds that

$$\begin{aligned}
&\tau^{U_k}(\beta/2) + 2\tau^{W_k}(\beta/2) \\
&= \left(18\overline{\mathrm{v}}_{f|\varphi}\mathrm{C}_\varphi + 138986\frac{L_{\beta/4}}{\beta}\frac{\sqrt{2k}}{n}\right) L_{\beta/4}^{1/2}\frac{\sqrt{\mathbf{v}_k}}{n} + 52\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} L_{\beta/4}\frac{\sqrt{\mathbf{m}_k}}{n} + 4\frac{\sqrt{\mathbf{m}_k}}{n}\frac{\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}}{\beta} \\
&\quad + \frac{1}{2}q_k^2(f - f_o) \\
&\leq \left(18\overline{\mathrm{v}}_{f|\varphi}\mathrm{C}_\varphi + 138986\frac{L_{\beta/4}}{\beta}\frac{\sqrt{2k}}{n}\right) L_{\beta/4}^{1/2}\frac{\sqrt{\mathbf{v}_k}}{n} + 56\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\frac{L_{\beta/4}}{\beta}\frac{\sqrt{\mathbf{m}_k}}{n} + \frac{1}{2}q_k^2(f - f_o).
\end{aligned} \tag{4.19}$$

Reformulating the separation condition given in Equation (4.16), yields that

$$q_k^2(f-f_o) \geq \tau + \left(18\mathrm{C}_\varphi \overline{\mathrm{v}}_{f|\varphi} + 138986\frac{L_{\beta/4}}{\beta}\frac{\sqrt{2k}}{n}\right) L_{\beta/4}^{1/2}\frac{\sqrt{\mathbf{v}_k}}{n} + 56\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\frac{L_{\beta/4}}{\beta}\frac{\sqrt{\mathbf{m}_k}}{n}$$
$$+\frac{1}{2}q_k^2(f-f_o).$$

Plugging Equation (4.19) into the last inequality, it follows that

$$q_k^2(f-f_o) \geq \tau + \tau^{U_k}(\beta/2) + 2\tau^{W_k}(\beta/2). \tag{4.20}$$

Thus, combining Equations (4.17), (4.18) and (4.20) the decomposition given in Equation (4.13) implies

$$\begin{aligned}
&\mathbb{P}_{f,\varphi}^n\left(\hat{q}_k^2 < \tau\right) = \mathbb{P}_{f,\varphi}^n\left(\{\hat{q}_k^2 < \tau\} \cap \Omega_U\right) + \mathbb{P}_{f,\varphi}^n\left(\{\hat{q}_k^2 < \tau\} \cap \Omega_U^c\right) \\
&\leq \mathbb{P}_{f,\varphi}^n(\Omega_U) + \mathbb{P}_{f,\varphi}^n\left(2W_k < \tau + \tau^{U_k}(\beta/2) - \hat{q}_k^2\right) \\
&\leq \frac{\beta}{2} + \mathbb{P}_{f,\varphi}^n(2W_k < -2\tau^{W_k}(\beta/2)) \\
&\leq \beta.
\end{aligned}$$

This completes the proof. □

4.3 Upper bound for the radius of testing

For $k \in \mathbb{N}$ and $\alpha \in (0,1)$, we use the test statistic $\hat{q}_k^2$ defined in Equation (4.7) and the critical value $\tau_k(\alpha)$ given in Equation (4.12) to define the test

$$\Delta_{k,\alpha} := \mathbb{1}_{\{\hat{q}_k^2 \geq \tau_k(\alpha)\}}. \tag{4.21}$$

From Proposition 4.2.2 follows immediately that the test $\Delta_{k,\alpha}$ is a level-α-test for all $k \in \mathbb{N}$. To analyze its power over the alternative, we consider the regularity class $\mathscr{F}^{\mathrm{s}|R}$ defined in Equation (4.3). Proposition 4.2.3 allows to characterize elements in $\mathscr{F}^{\mathrm{s}|R}$ for which $\Delta_{k,\alpha}$ is powerful. Exploiting these results, we derive in this section an upper bound for the radius of testing of $\Delta_{k,\alpha}$ in terms of $\mathbf{v}_k$ and $\mathbf{m}_k$ defined in Equations (4.8) and (4.9), respectively, and the regularity parameter s. That is, define

$$\rho_{k,\mathrm{s}}^2 := \rho_{k,\mathrm{s}}^2(n) := \frac{\mathrm{w}^2(k)}{\mathrm{s}^2(k)} \vee \frac{1}{n}\left(\sqrt{\mathbf{v}_k} \vee \sqrt{\mathbf{m}_k}\right).$$

For the result on the upper bound of the radius of testing, we need a higher moment condition for the elements of the alternative to apply Proposition 4.2.3 and get a uniform bound. More precisely, we introduce for $\mathrm{v} \in \mathbb{R}$, $\mathrm{v} \geq 1$ the set

$$\mathscr{D}^{\mathrm{v}} := \{h \in \mathscr{D} \cap \mathbb{L}_+^1(\mathrm{x}^{4(c-1)}) : \overline{\mathrm{v}}_{h|\varphi} \leq \mathrm{v}\}.$$

The constant $\overline{\mathrm{v}}_{h|\varphi}$ is defined in Equation (4.11) and finite for $h \in \mathbb{L}_+^1(\mathrm{x}^{4(c-1)})$. The following result gives a bound on the maximal risk of the proposed test.

Proposition 4.3.1 (Upper bound for the radius of testing of $\Delta_{k,\gamma/2}$):
Under Assumptions 4.1.1 to 4.1.3, for $\gamma \in (0,1)$ and $\mathrm{d} \in \mathbb{R}_+$ define

$$A_\gamma^2 := R^2 + 140\frac{L_{\gamma/8}}{\gamma}\mathrm{C}_\varphi \overline{\mathrm{v}}_{f_o|\varphi} + 260\frac{L_{\gamma/8}}{\gamma}\mathrm{C}_\varphi\,\mathrm{v} + 833934\frac{L_{\gamma/8}^{3/2}}{\gamma}\,\mathrm{d} \tag{4.22}$$

and

$$\eta_k^2 := \eta_k^2(n) := 1 \vee \frac{\sqrt{2k}}{\mathrm{d}\, n}.$$

For all $A \geq A_\gamma$ and $n, k \in \mathbb{N}$ with $n \geq 2$ we have

$$\mathcal{R}\left(\Delta_{k,\gamma/2} | \mathscr{D}^{\mathrm{v}}, \mathscr{F}^{\mathrm{s}|R}, A^2\rho_{k,\mathrm{s}}^2 \cdot \eta_k^2\right) \leq \gamma.$$

Proof of Proposition 4.3.1. The result follows from Proposition 4.2.2 and Proposition 4.2.3 with $\alpha = \beta = \gamma/2$, $\tau = \tau_k(\alpha)$ and the definition of the maximal risk

$$\begin{aligned}\mathcal{R}\left(\Delta_{k,\gamma/2} | \mathscr{D}^{\mathrm{v}}, \mathscr{F}^{\mathrm{s}|R}, A^2\rho_{k,\mathrm{s}}^2 \cdot \eta_k^2\right) &= \mathbb{P}^n_{\varphi,f_o}\left(\Delta_{k,\gamma/2} = 1\right) \\ &\quad + \sup_{f \in \mathscr{D}^{\mathrm{v}} \cap (f_o + \mathscr{F}^{\mathrm{s}|R} \cap \mathscr{F}_{\geq A^2\rho_{k,\mathrm{s}}^2 \cdot \eta_k^2})} \mathbb{P}^n_{\varphi,f}\left(\Delta_{k,\gamma/2} = 0\right) \\ &\leq \frac{\gamma}{2} + \frac{\gamma}{2} = \gamma.\end{aligned}$$

To conclude the proof, we need to show that the assumptions of both propositions are satisfied. For Proposition 4.2.2 this is the case. It remains to check the assumptions for Proposition 4.2.3, that is, verify Equation (4.16) for elements in the alternative $f \in \mathscr{D}^{\mathrm{v}} \cap (f_o + \mathscr{F}^{\mathrm{s}|R} \cap \mathscr{F}_{\geq A^2\rho_{k,\mathrm{s}}^2 \cdot \eta_k^2})$. More precisely, it remains to verify

$$\begin{aligned}q_k^2(f - f_o) \geq 2\tau_k(\gamma/2) + 2\left(18\mathrm{C}_\varphi \overline{\mathrm{v}}_{f|\varphi} + 2 \cdot 138986\frac{L_{\gamma/8}}{\gamma}\frac{\sqrt{2k}}{n}\right) L_{\gamma/8}^{1/2}\frac{\sqrt{\mathbf{v}_k}}{n} \\ + 2 \cdot 112\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\frac{L_{\gamma/8}}{\gamma}\frac{\sqrt{\mathbf{m}_k}}{n}.\end{aligned} \tag{4.23}$$

Indeed, with Assumption 4.1.2 and $f - f_o \in \mathscr{F}^{\mathrm{s}|R}$, we bound the bias under the alternative. More precisely, it holds that

$$\begin{aligned}q^2(f - f_o) - q_k^2(f - f_o) &= \|\mathbb{1}_{[-k,k]^c}(\mathcal{M}_c[f] - \mathcal{M}_c[f_o])\|^2_{\mathbb{L}^2(\mathrm{w}^2)} \\ &= \|\mathbb{1}_{[-k,k]^c}(\mathcal{M}_c[f] - \mathcal{M}_c[f_o])\mathrm{s}/\mathrm{s}\|^2_{\mathbb{L}^2(\mathrm{w}^2)} \\ &\leq \|\mathbb{1}_{[-k,k]^c}/\mathrm{s}\|^2_{\mathbb{L}^\infty(\mathrm{w})}\|\mathrm{s}(\mathcal{M}_c[f] - \mathcal{M}_c[f_o])\|^2_{\mathbb{L}^2(\mathrm{w}^2)} \\ &\leq \frac{\mathrm{w}^2(k)}{\mathrm{s}^2(k)}R^2.\end{aligned} \tag{4.24}$$

Consequently, we have for $f - f_o \in \mathscr{F}^{\mathrm{s}|R} \cap \mathscr{F}_{\geq A^2\rho_{k,\mathrm{s}}^2 \cdot \eta_k^2}$ have that

$$q_k^2(f - f_o) \geq q^2(f - f_o) - \frac{\mathrm{w}^2(k)}{\mathrm{s}^2(k)}R^2 \geq A^2\rho_{k,\mathrm{s}}^2 \cdot \eta_k^2 - \frac{\mathrm{w}^2(k)}{\mathrm{s}^2(k)}R^2. \tag{4.25}$$

Let us upper bound the quantile $\tau_k(\gamma/2)$. Since $L_\alpha = 1 - \log \alpha \geq 1$ for all $\alpha \in (0,1)$ and L_α is increasing in α, we also have that $L_{\gamma/4}, L_{\gamma/4}^{1/2} \leq L_{\gamma/8}$. Thus, it follows

$$L_{\gamma/4} L_{\gamma/4}^{1/2} \leq L_{\gamma/8}^{3/2}.$$

Additionally, because $\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} \leq \mathrm{C}_\varphi^2 \overline{\mathrm{v}}_{f|\varphi}$ we obtain for the critical value

$$\begin{aligned}
\tau_k(\gamma/2) &= \left(18\mathrm{C}_\varphi \overline{\mathrm{v}}_{f_o|\varphi} + 138986 \frac{\sqrt{2k}}{n} \frac{L_{\gamma/4}}{\gamma} \right) L_{\gamma/4}^{1/2} \frac{\sqrt{\mathbf{v}_k}}{n} + 52 \mathrm{v}_{f_o|\varphi} \mathrm{C}_\varphi L_{\gamma/4} \frac{\sqrt{\mathbf{m}_k}}{n} \\
&\leq \left(18 \frac{L_{\gamma/8}}{\gamma} \mathrm{C}_\varphi \overline{\mathrm{v}}_{f_o|\varphi} + 138986 \frac{\sqrt{2k}}{n} \frac{L_{\gamma/8}^{3/2}}{\gamma} \right) \frac{\sqrt{\mathbf{v}_k}}{n} + 52 \mathrm{v}_{f_o|\varphi} \mathrm{C}_\varphi \frac{L_{\gamma/8}}{\gamma} \frac{\sqrt{\mathbf{m}_k}}{n} \\
&\leq \left(70 \overline{\mathrm{v}}_{f_o|\varphi} \mathrm{C}_\varphi \frac{L_{\gamma/8}}{\gamma} + 138986 \frac{\sqrt{2k}}{n} \frac{L_{\gamma/8}^{3/2}}{\gamma} \right) \left(\frac{\sqrt{\mathbf{v}_k}}{n} \vee \frac{\sqrt{\mathbf{m}_k}}{n} \right).
\end{aligned} \tag{4.26}$$

Using $A \geq A_\gamma$ for A_γ defined in Equation (4.22) and $\mathrm{d}\, \eta_k^2 \geq \frac{\sqrt{2k}}{n}$, we get for Equation (4.25) that

$$\begin{aligned}
q_k^2(f - f_o) &\geq \left(R^2 + 140 \frac{L_{\gamma/8}}{\gamma} \mathrm{C}_\varphi \overline{\mathrm{v}}_{f_o|\varphi} + 260 \frac{L_{\gamma/8}}{\gamma} \mathrm{C}_\varphi\, \mathrm{v} + 833934 \frac{L_{\gamma/8}^{3/2}}{\gamma} \eta_k^2\, \mathrm{d} \right) \\
&\quad \cdot \left(\frac{\mathrm{w}^2(k)}{\mathrm{s}^2(k)} \vee \frac{1}{n} \left(\sqrt{\mathbf{v}_k} \vee \sqrt{\mathbf{m}_k} \right) \right) - \frac{\mathrm{w}^2(k)}{\mathrm{s}^2(k)} R^2 \\
&\geq \left((2 \cdot 70) \frac{L_{\gamma/8}}{\gamma} \mathrm{C}_\varphi \overline{\mathrm{v}}_{f_o|\varphi} + (2 \cdot 18 + 2 \cdot 112) \frac{L_{\gamma/8}}{\gamma} \mathrm{C}_\varphi \overline{\mathrm{v}}_{f|\varphi} + (2+4) 138986 \frac{L_{\gamma/8}^{3/2}}{\gamma} \frac{\sqrt{2k}}{n} \right) \\
&\quad \cdot \frac{1}{n} \left(\sqrt{\mathbf{v}_k} \vee \sqrt{\mathbf{m}_k} \right).
\end{aligned}$$

Combining the calculation with the bound derived in Equation (4.26) for the quantile $\tau_k(\gamma/2)$, we finally obtain

$$\begin{aligned}
q_k^2(f - f_o) &\geq 2\tau_k(\gamma/2) + 2 \left(18 \mathrm{C}_\varphi \overline{\mathrm{v}}_{f|\varphi} + 2 \cdot 138986 \frac{L_{\gamma/8}}{\gamma} \frac{\sqrt{2k}}{n} \right) L_{\gamma/8}^{1/2} \frac{\sqrt{\mathbf{v}_k}}{n} \\
&\quad + 2 \cdot 112 \mathrm{C}_\varphi \mathrm{v}_{f|\varphi} \frac{L_{\gamma/8}}{\gamma} \frac{\sqrt{\mathbf{m}_k}}{n},
\end{aligned}$$

which shows the separation condition given in Equation (4.23) and completes the proof. □

Let us introduce a dimension parameter that realizes an optimal bias-variance trade-off and the corresponding radius:

$$k_{\mathrm{s}}^* := k_{\mathrm{s}}^*(n) := \arg\min_{k \in \mathbb{N}} \rho_{k,\mathrm{s}}^2(n) \quad \text{and} \quad \rho_{*,\mathrm{s}}^2 := \rho_{*,\mathrm{s}}^2(n) = \min_{k \in \mathbb{N}} \rho_{k,\mathrm{s}}^2(n). \tag{4.27}$$

The next result follows immediately from Proposition 4.3.1.

Corollary 4.3.2:
Under the assumptions of Proposition 4.3.1 let $\gamma \in (0,1)$ and A_γ as in Equation (4.22). For all $A \geq A_\gamma$ and $n \geq 2$ we have

$$\mathcal{R}\left(\Delta_{k_{\mathrm{s}}^*,\gamma/2}|\mathscr{D}^{\mathrm{v}}, \mathscr{F}^{\mathrm{s}|R}, A^2\rho_{*,\mathrm{s}}^2 \cdot \eta_{k_{\mathrm{s}}^*}^2\right) \leq \gamma.$$

For two real-valued functions $h_1, h_2\colon \mathbb{R} \to \mathbb{R}$ we write $h_1(t) \sim h_2(t)$ if there exist constants $C, \tilde{C}$ such that for all $t \in \mathbb{R}$ we have $h_1(t) \leq C h_2(t)$ and $h_2(t) \leq \tilde{C} h_1(t)$. We then call h_2 the order of h_1.

We determine the order of the radius of testing

$$\rho_*^2(n) = \min_{k\in\mathbb{N}}\left(\frac{\mathrm{w}^2(k)}{\mathrm{s}^2(k)} \vee \frac{1}{n}\left(\sqrt{\mathbf{v}_k} \vee \sqrt{\mathbf{m}_k}\right)\right)$$

for different choices of regularity parameter s, density function w^2 and order of the Mellin transform of the error density, which determines the order of $\mathbf{v}_k$ and $\mathbf{m}_k$. We consider the standard cases, as already discussed in Chapter 3.

The class $\mathscr{F}^{\mathrm{s}|R}$ defined in Equation (4.3) covers the usual assumptions on the regularity of the unknown density f, i.e., ordinary and super smooth densities. More precisely, in the *ordinary smooth case* the weight function s is of polynomial order, i.e., $\mathrm{s}(t) \sim (1+t^2)^{s/2}$ for $t \in \mathbb{R}$ and some $s \in \mathbb{R}_+$. In the *super smooth* case s is assumed to be exponentially increasing, i.e., $\mathrm{s}(t) \sim \exp(|t|^s)$ for $t \in \mathbb{R}$ some $s \in \mathbb{R}_+$. In the ordinary smooth case, $\mathscr{F}^{\mathrm{s}|R}$ corresponds to the Mellin-Sobolev space, see Definition 2.3.9, Brenner Miguel (2023). Concerning Assumption 4.1.2 (ii), boundedness of w / s ensures that the quadratic functional is finite on the nonparametric class of functions $\mathscr{F}^{\mathrm{s}|R}$. Convergence against zero gives the rate for the bias term.

For the density function, we consider the case that $\mathrm{w}(t) \sim t^a$ for $a \in \mathbb{R}$. This includes Examples 1.4.1 to 1.4.3 with the following choices of a.

(i) In the case $a = 0$ we have $\mathrm{w}(t) = 1$ for all $t \in \mathbb{R}$. Recall that in this case q^2 equals a (weighted) $\mathbb{L}_+^2$-norm of the density f itself.

(ii) The case $a = -1$ covers quadratic functional estimation of the survival function S_f of X, more precisely, we have $\mathrm{w}^2(t) = \frac{1}{(c-1)^2+4\pi t^2}$ for all $t \in \mathbb{R}$.

(iii) The case $a = \beta \in \mathbb{N}$ covers quadratic functionals of derivatives $D^\beta[f]$.

Note that, in case of ordinary smoothness of the unknown density f, Assumption 3.3.1 (ii) imposes the condition $s > a$ on the parameters. Otherwise all the examples of parameter a satisfy Assumption 4.1.2 for t large enough for both the case of ordinary smoothness and super smoothness.

As in the previous chapter, we consider the following examples for the behavior of the Mellin transform of the error density φ. We distinguish again between two cases. We either assume for some decay parameter $\sigma \in \mathbb{R}_+$ its *(ordinary) smoothness*, i.e., $|\mathcal{M}_c[\varphi](t)| \sim (1+t^2)^{-\sigma/2}$; or its *super smoothness*, i.e., $|\mathcal{M}_c[\varphi](t)| \sim \exp(-|t|^\sigma)$.

We note that $\rho_{k,\mathrm{s}}^2(n)$ corresponds to the term $R_n^2(k)$ appearing in the upper bound of the mean squared error of the quadratic functional estimation problem and the discussion on the rates for the standard regularity assumptions, for more details see Section 3.3. Consequently, the order of $R_n(k_*)$ corresponds to the order of $\sqrt{\rho_{k_{\mathrm{s}}^*,\mathrm{s}}^2(n)}$. In Table 4.1 we give the corresponding rates.

For the case of both densities to be ordinary smooth (first line), we assume that $\sigma + a > -\frac{1}{4}$. Let us remark that in the case of $\sigma + a = -\frac{1}{4}$ the term $\sqrt{\mathbf{v}_k} \vee \sqrt{\mathbf{m}_k} = \sqrt{\mathbf{v}_k}$ is of order $\sqrt{\log k}$ and the corresponding rate of testing is of order $\sqrt{\log n}/n$. If the density function w is even of smaller order, i.e., $\sigma + a < -\frac{1}{4}$, it holds that $\sqrt{\mathbf{v}_k} \vee \sqrt{\mathbf{m}_k} \sim 1$ and we achieve a parametric rate n^{-1}.

Note that in the severely ill-posed model with ordinary smoothness (second line), the optimal parameter k_{s}^* does not depend on the regularity parameter s. Hence, in this case the test $\Delta_{k_{\mathrm{s}}^*}$ is automatically adaptive. Note also that in all three examples the factor

$$\eta_{k_{\mathrm{s}}^*}^2 = 1 \vee \frac{\sqrt{2k_{\mathrm{s}}^*}}{\mathrm{d}\,n}$$

is of order one and, thus, the order of the radius of testing is indeed equal to the order of $\rho_{*,\mathrm{s}}^2(n)$.

$\mathrm{s}(t)$	$\lvert\mathcal{M}_c[\varphi](t)\rvert$	k_{s}^*	$\rho_{*,\mathrm{s}}^2(n)$
$(1+t^2)^{-\frac{s}{2}}$	$(1+t^2)^{-\frac{\sigma}{2}}$	$n^{\frac{2}{4s+4\sigma+1}}$	$n^{-\frac{4(s-a)}{4s+4\sigma+1}}$
$(1+t^2)^{-\frac{s}{2}}$	$\exp(-\lvert t\rvert^{\sigma})$	$(\log n)^{\frac{1}{\sigma}}$	$(\log n)^{-\frac{2(s-a)}{\sigma}}$
$\exp(-\lvert t\rvert^{s})$	$(1+t^2)^{-\frac{\sigma}{2}}$	$(\log n)^{\frac{1}{s}}$	$\frac{1}{n}(\log n)^{\frac{2(\sigma+a)+1/2}{s}}$

Table 4.1: Order of the radius of testing for $\mathrm{w}(t) \sim t^a$ for $t \in \mathbb{R}$, $a \in \mathbb{R}$.

4.4 Testing radius of a max-test

The test $\Delta_{k,\gamma/2}$ in Proposition 4.3.1 relies on the dimension parameter k. The optimal choice of k in turn depends on the smoothness class $\mathscr{F}^{\mathrm{s}\,|R}$, see Table 4.1. Ideally, we want our testing procedure to be adaptive, i.e., assumption-free, with respect to the alternative class $\mathscr{F}^{\mathrm{s}\,|R}$. It should perform optimally for a wide range of alternatives. In this section, we therefore propose an data-driven testing procedure by aggregating the tests derived in the last section over various dimension parameters k. In Section 4.4.1, we first introduce the testing procedure to derive in Section 4.4.2 an upper bound for the radius of testing.

4.4.1 Data-driven procedure via Bonferroni aggregation

We first recapitulate the main ideas of the Bonferroni aggregation. Let $\mathcal{K} \subseteq \mathbb{N}$ be a finite collection of dimension parameters. Recall that the following collection of tests are tests of level $\alpha/|\mathcal{K}| \in (0,1)$

$$(\Delta_{k,\alpha/|\mathcal{K}|})_{k\in\mathcal{K}} := (\mathbb{1}_{\{\hat{q}_k^2 \geq \tau_k(\alpha/|\mathcal{K}|)\}})_{k\in\mathcal{K}}.$$

We consider the max-test

$$\Delta_{\mathcal{K},\alpha} := \mathbb{1}_{\{\zeta_{\mathcal{K},\alpha}>0\}} \quad \text{with} \quad \zeta_{\mathcal{K},\alpha} = \max_{k\in\mathcal{K}} \Delta_{k,\alpha/|\mathcal{K}|}$$

i.e. the test rejects the null hypothesis as soon as one of the tests in the collection does. Under the null hypothesis, we bound the type I error probability of the max-test by the sum of the error probabilities of the individual tests

$$\mathbb{P}^n_{\varphi,f_o}(\Delta_{\mathcal{K},\alpha} = 1) = \mathbb{P}^n_{\varphi,f_o}(\zeta_{\mathcal{K},\alpha} > 0) \leq \sum_{k\in\mathcal{K}} \mathbb{P}^n_{\varphi,f_o}(\Delta_{k,\alpha/|\mathcal{K}|} = 1) \leq \sum_{k\in\mathcal{K}} \frac{\alpha}{|\mathcal{K}|} = \alpha. \tag{4.28}$$

Hence, $\Delta_{\mathcal{K},\alpha}$ is a level-α-test applying Proposition 4.2.2. Under the alternative the type II error probability is bounded by the minimum of the error probabilities of the individual tests, i.e.

$$\mathbb{P}^n_{\varphi,f}(\Delta_{\mathcal{K},\alpha} = 0) = \mathbb{P}^n_{\varphi,f}(\zeta_{\mathcal{K},\alpha} \leq 0) \leq \min_{k\in\mathcal{K}} \mathbb{P}^n_{\varphi,f}(\Delta_{k,\alpha/|\mathcal{K}|} = 0). \tag{4.29}$$

Therefore, $\Delta_{\mathcal{K},\alpha}$ has the maximal power achievable by a test in the collection. The bounds Equation (4.28) and Equation (4.29) have opposing effects on the choice of the collection $\mathcal{K}$. On the one hand, it should be as small as possible to keep the type I error probability small. On the other hand, it must be large enough to approximate an optimal dimension parameter k^*_{s} for a wide range of regularity parameters s that we want to adapt to. In this work we consider a classical Bonferroni choice of an error level $\alpha/|\mathcal{K}|$, For other aggregation choices, e.h. a Monte Carlo quantile and a Monte Carlo threshold method we refer to Baraud et al. (2003) and Fromont and Laurent (2006). Although the Bonferroni choice is a more conservative method, it is often shown its optimality, which is also shared with the other methods.

Applying the Bonferroni rule to the level $\alpha/|\mathcal{K}| \in (0,1)$, the critical value in Equation (4.12) writes

$$\tau_k(\alpha/|\mathcal{K}|) := \left(18\mathrm{C}_\varphi \overline{\mathrm{v}}_{f_o|\varphi} + 69493\frac{\sqrt{2k}}{n}\frac{L_{\alpha/2|\mathcal{K}|}|\mathcal{K}|}{\alpha}\right) L^{1/2}_{\alpha/2|\mathcal{K}|}\frac{\sqrt{\mathbf{v}_k}}{n} + 52\mathrm{v}_{f_o|\varphi}\mathrm{C}_\varphi L_{\alpha/2|\mathcal{K}|}\frac{\sqrt{\mathbf{m}_k}}{n}$$

We observe that the critical value and hence the separation condition Equation (4.16) in Proposition 4.2.3 increases at least by a factor $|\mathcal{K}|$ leading possibly to a deterioration of the radius of testing by this factor. This is known to be suboptimal. To adapt the critical value accordingly, we impose in this section a stronger moment condition under the null hypothesis. That is, instead of Assumption 4.2.1, we will assume the following in the results of this section.

Assumption 4.4.1:
For $p \geq 2$ let $\varphi \in \mathbb{L}^\infty_+(\mathrm{x}^{2c-1}) \cap \mathbb{L}^1_+(\mathrm{x}^{2p(c-1)})$ and $f_o \in \mathbb{L}^1_+(\mathrm{x}^{2p(c-1)})$.

Under Assumption 4.4.1 , it follows that $f_o \circledast \varphi \in \mathbb{L}^1_+(\mathrm{x}^{2p(c-1)})$ by Property 1.2.3 (iii) and we write shortly

$$\mathrm{v}_{f_o|\varphi}(p) := \|f_o \circledast \varphi\|_{\mathbb{L}^1_+(\mathrm{x}^{2p(c-1)})} \vee 1.$$

Finally, we introduce an updated critical value

$$\tau_{k|\alpha} := \left(18\mathrm{C}_\varphi \mathrm{v}_{f_o|\varphi}(p) + 69493\frac{\sqrt{2k}|\mathcal{K}|^{1/(p-1)}}{n}\frac{L^2_{\alpha/2|\mathcal{K}|}}{\alpha}\right) L^{1/2}_{\alpha/2|\mathcal{K}|}\frac{\sqrt{\mathbf{v}_k}}{n} + 52\mathrm{v}_{f_o|\varphi}\mathrm{C}_\varphi L_{\alpha/2|\mathcal{K}|}\frac{\sqrt{\mathbf{m}_k}}{n}. \tag{4.30}$$

Before giving an upper bound of the radius of testing of the max-test, we generalize Proposition 4.2.2 to the updated critical value.

Proposition 4.4.2 (Bound for the quantiles of $\hat{q}^2_k$ under the null hypothesis)**:**
Let Assumptions 4.1.1, 4.4.1 and 4.1.3 be satisfied and let $\alpha \in (0,1)$, $n \geq 2$ and $k \in \mathbb{N}$. Consider the estimator $\hat{q}^2_k$ and the threshold $\tau_{k|\alpha}$ defined in (4.7) and (4.30), respectively. Under the null hypothesis we have that

$$\mathbb{P}^n_{\varphi,f_o}(\hat{q}^2_k \geq \tau_{k|\alpha}) \leq \frac{\alpha}{|\mathcal{K}|}.$$

Proof of Proposition 4.4.2. The proof follows along the same lines as the proof of Proposition 4.2.2. If $f = f_o$ and, hence $g = g_o$ and $\mathcal{M}_c[g] = \mathcal{M}_c[g_o]$, the decomposition (4.13) simplifies to $\hat{q}_k^2 = U_k$. Due to Assumption 4.4.1, we have that $f_o, \varphi \in \mathbb{L}_+^1(\mathrm{x}^{2p(c-1)})$ and the assumptions of Lemma 4.7.4 are satisfied. Moreover, with $\gamma = \alpha/|\mathcal{K}|$, using the notations in (4.8) and (4.9) and the fact that for $p \geq 2$ it holds that $\alpha^{-1/(p-1)} \leq \alpha^{-1}$ and $L_{\alpha/2|\mathcal{K}|}^{2-1/(p-1)} \leq L_{\alpha/2|\mathcal{K}|}^2$, we have that

$$\tau^{U_k}(\alpha/|\mathcal{K}|) = \left(18\mathrm{v}_{f_o|\varphi}(p)\mathrm{C}_\varphi + 69493\frac{\sqrt{2k}|\mathcal{K}|^{1/(p-1)}}{n\alpha^{1/(p-1)}}L_{\alpha/2|\mathcal{K}|}^{2-1/(p-1)}\right)\frac{\sqrt{\mathbf{v}_k}}{n}L_{\alpha/2|\mathcal{K}|}^{1/2}$$
$$+ 52\mathrm{C}_\varphi \mathrm{v}_{f_o|\varphi}\frac{\sqrt{\mathbf{m}_k}}{n}L_{\alpha/2|\mathcal{K}|} \leq \tau_{k|\alpha}.$$

Thus, we obtain immediately from Lemma 4.7.4 that

$$\mathbb{P}_{\varphi,f_o}^n(\hat{q}_k^2 \geq \tau_{k|\alpha}) \leq \mathbb{P}_{\varphi,f_o}^n(\hat{q}_k^2 \geq \tau^{U_k}(\alpha/|\mathcal{K}|)) \leq \frac{\alpha}{|\mathcal{K}|}.$$

This concludes the proof. □

Due to Proposition 4.4.2, for any finite collection $\mathcal{K} \subset \mathbb{N}$ and any $\alpha \in (0,1)$ we have a collection of level-$\frac{\alpha}{|\mathcal{K}|}$-tests given by

$$\left(\Delta_{k|\alpha} := \mathbb{1}_{\{\hat{q}_k^2 \geq \tau_{k|\alpha}\}}\right)_{k\in\mathcal{K}}$$

and obtain the max-test

$$\Delta_{\mathcal{K}|\alpha} := \mathbb{1}_{\{\zeta_{\mathcal{K}|\alpha} > 0\}} \quad \text{with} \quad \zeta_{\mathcal{K}|\alpha} = \max_{k\in\mathcal{K}} \Delta_{k|\alpha}. \tag{4.31}$$

In the next section, we derive an upper bound for its radius of testing.

4.4.2 Testing radius of the max-test

Denote by $\mathcal{S}$ the set of functions satisfying Assumption 4.1.2. The set $\mathcal{S}$ characterizes the collection of alternatives $\{\mathscr{F}^{\mathrm{s}|R} : \mathrm{s} \in \mathcal{S}\}$ for which we analyze the power of our testing procedure simultaneously. The max-test only aggregates over a finite set $\mathcal{K} \subseteq \mathbb{N}$. We define the minimal achievable radius of testing for each $\mathrm{s} \in \mathcal{S}$ over the set $\mathcal{K}$ for each $x \in \mathbb{R}, x \geq 1$ as

$$\rho_{\mathcal{K},s}^2(x) := \min_{k\in\mathcal{K}} \rho_{k,s}^2(x) \quad \text{with} \quad \rho_{k,s}^2(x) := \frac{\mathrm{w}^2(k)}{\mathrm{s}^2(k)} \vee \frac{1}{x}\left(\sqrt{\mathbf{v}_k} \vee \sqrt{\mathbf{m}_k}\right).$$

Since ρ_*^2 is defined as the minimum taken over $\mathbb{N}$ instead of $\mathcal{K}$ for each $n \in \mathbb{N}$ we always have $\rho_{\mathcal{K},s}^2(n) \geq \rho_{\mathbb{N},s}^2(n) = \rho_*^2$ and evidently $\rho_{\mathcal{K},s}^2(n) = \rho_{\mathbb{N},s}^2(n)$ whenever $k_{\mathrm{s}}^* \in \mathcal{K}$. Let us introduce the adaptive factor $\delta_{\mathcal{K}} := (1 + \log|\mathcal{K}|)^{-1/2} \in (0,1)$. For each $\mathrm{s} \in \mathcal{S}$, $k \in \mathcal{K}$ and $x \in \mathbb{R}, x \geq 1$ we set

$$\mathrm{r}_{k,\mathrm{s}}^2(x) := \frac{\mathrm{w}^2(k)}{\mathrm{s}^2(k)} \vee \frac{1}{x}\left(\delta_{\mathcal{K}}^{-1}\sqrt{\mathbf{v}_k} \vee \delta_{\mathcal{K}}^{-2}\sqrt{\mathbf{m}_k}\right)$$
$$k_{\mathcal{K},\mathrm{s}}(x) := \arg\min_{k\in\mathcal{K}} \mathrm{r}_{k,\mathrm{s}}^2(x) \tag{4.32}$$
$$\mathrm{r}_{\mathcal{K},\mathrm{s}}^2(x) := \min_{k\in\mathcal{K}} \mathrm{r}_{k,\mathrm{s}}^2(x). \tag{4.33}$$

In the next proposition we show for each $\mathrm{s} \in \mathcal{S}$ that both the type I and maximal type II error probability of a max-test $\Delta_{\mathcal{K}|\gamma/2}$ are bounded by $\gamma/2$ and thus their sum by γ, by applying Propositions 4.4.2 and 4.2.3 with $\alpha = \beta = \gamma/2$.

Proposition 4.4.3 (Uniform radius of testing over regularity class $\mathcal{S}$)**:**
Under Assumptions 4.1.1, 4.4.1 and 4.1.3 and let $\gamma \in (0,1)$ and $\mathrm{d} \in \mathbb{R}, \mathrm{d} \geq 1$ define

$$A_\gamma^2 := R^2 + 140\frac{L_{\gamma/8}}{\gamma}\mathrm{C}_\varphi \mathrm{v}_{f_o|\varphi}(p) + 260\frac{L_{\gamma/8}}{\gamma}\mathrm{C}_\varphi\, \mathrm{v} + 833934\frac{L_{\gamma/8}^{5/2}}{\gamma}\,\mathrm{d}\,.$$

For all $A \geq A_\gamma$, $n \geq 2$, $\mathrm{s} \in \mathcal{S}$ with $k_{\mathcal{K},\mathrm{s}}(n) \in \mathcal{K} \subseteq \mathbb{N}$ as defined in Equation (4.32) and

$$\eta_{\mathcal{K}|n}^2 := 1 \vee \frac{\sqrt{2k_{\mathcal{K},\mathrm{s}}(n)}|\mathcal{K}|^{1/(p-1)}}{\mathrm{d}\, n\delta_{\mathcal{K}}^4}, \tag{4.34}$$

we have

$$\mathcal{R}\left(\Delta_{\mathcal{K}|\gamma/2}|\mathscr{D}^{\mathrm{v}}, \mathscr{F}^{\mathrm{s}|R}, A^2\, \mathrm{r}_{\mathcal{K},\mathrm{s}}^2(n)\cdot\eta_{\mathcal{K}|k}^2\right) \leq \gamma.$$

Proof of Proposition 4.4.3. Under the null hypothesis the claim follows for each $\mathrm{s} \in \mathcal{S}$ from Equation (4.28) together with Proposition 4.4.2, that is,

$$\mathbb{P}_{\varphi,f_o}^n\left(\Delta_{k|\gamma/2} = 1\right) = \mathbb{P}_{\varphi,f_o}^n\left(\hat{q}_k^2 \geq \tau_{k|\gamma/2}\right) \leq \frac{\gamma}{2|\mathcal{K}|}$$

for all $k \in \mathcal{K}$, and $\sum_{k\in\mathcal{K}} \frac{\gamma}{2|\mathcal{K}|} = \gamma/2$. Thus, it holds

$$\mathbb{P}_{\varphi,f_o}^n\left(\Delta_{\mathcal{K}|\gamma/2} = 1\right) \leq \frac{\gamma}{2}$$

For $f \in \mathscr{D}^{\mathrm{v}}$ satisfying separation condition Equation (4.16) for $\tau_{k|\gamma/2}$ the assumptions of Proposition 4.2.3 are fulfilled, and it holds with Equation (4.29) that

$$\mathbb{P}_{\varphi,f}^n\left(\Delta_{\mathcal{K}|\gamma/2} = 0\right) \leq \min_{k\in\mathcal{K}} \mathbb{P}_{\varphi,f}^n\left(\Delta_{k|\gamma/2} = 0\right) \leq \frac{\gamma}{2|\mathcal{K}|} \leq \frac{\gamma}{2}.$$

It remains to show that elements of the alternative satisfy the separation condition Equation (4.16) for $\tau_{k|\gamma/2}$ for some $k \in \mathcal{K}$. Let $f - f_o \in \mathscr{F}^{\mathrm{s}|R}$ satisfy $q^2(f-f_o) \geq A^2\,\mathrm{r}_{\mathcal{K},\mathrm{s}}^2(n)\cdot\eta_{\mathcal{K}|k}^2$ for some $A^2 \geq A_\gamma^2$. Further, set $k^\circ := k_{\mathcal{K},\mathrm{s}}(n) \in \mathcal{K}$ as defined in Equation (4.32). The separation condition Equation (4.16) then states

$$\begin{aligned} q_{k^\circ}^2(f-f_o) \geq 2\tau_{k^\circ|\gamma/2} + 2\left(18\mathrm{C}_\varphi\overline{\mathrm{v}}_{f|\varphi} + 2\cdot 138986\frac{L_{\gamma/8}}{\gamma}\frac{\sqrt{2k^\circ}}{n}\right)L_{\gamma/8}^{1/2}\frac{\sqrt{\mathbf{v}_{k^\circ}}}{n} \\ + 2\cdot 112\mathrm{C}_\varphi\mathrm{v}_{f|\varphi}\frac{L_{\gamma/8}}{\gamma}\frac{\sqrt{\mathbf{m}_{k^\circ}}}{n} \end{aligned} \tag{4.35}$$

with $\tau_{k^\circ|\gamma/2}$ as in Equation (4.30). Indeed, we have analogously to Equation (4.24)

$$q^2(f-f_o) - q_{k^\circ}^2(f-f_o) \leq \frac{\mathrm{w}^2(k^\circ)}{\mathrm{s}^2(k^\circ)}R^2$$

and, thus, since $f - f_o \in \mathscr{F}^{\mathrm{s}|R}$

$$q_{k^\circ}^2(f-f_o) \geq q^2(f-f_o) - \mathrm{s}^2(k^\circ)\,\mathrm{w}^2(k^\circ)R^2 \geq A_\gamma^2 r_{k^\circ,\mathrm{s}}^2\cdot\eta_{k^\circ}^2 - \frac{\mathrm{w}^2(k^\circ)}{\mathrm{s}^2(k^\circ)}R^2. \tag{4.36}$$

Using $L_{\gamma/4} \geq 1$ it holds

$$L_{\gamma/(4|\mathcal{K}|)} = 1 - \log(\gamma/(4|\mathcal{K}|)) = 1 - \log(\gamma/4) + \log(|\mathcal{K}|)$$
$$\leq (1 - \log(\gamma/4))(1 + \log(|\mathcal{K}|)) = L_{\gamma/4}\delta_{\mathcal{K}}^{-2}.$$

In addition, due to $p \geq 2$ it holds $2^{1/(p-1)} \leq 2$,

$$\frac{L_{\gamma/4}^2 L_{\gamma/4}^{1/2}}{\gamma^{1/(p-1)}} \leq \frac{L_{\gamma/8}^{5/2}}{\gamma}.$$

Also, we have $\mathrm{v}_{f_o|\varphi} \leq \mathrm{v}_{f_o|\varphi}(p)$. Thus, we obtain for the critical value

$$\begin{aligned}
\tau_{k^\circ|\gamma/2} = &\left(18\mathrm{C}_\varphi \mathrm{v}_{f_o|\varphi}(p) + 2\cdot 69493\frac{\sqrt{2k^\circ}|\mathcal{K}|^{1/(p-1)}}{n}\frac{L^2_{\frac{\gamma}{4|\mathcal{K}|}}}{\gamma}\right) L^{1/2}_{\frac{\gamma}{4|\mathcal{K}|}}\frac{\sqrt{\mathbf{v}_{k^\circ}}}{n} \\
&+ 52\mathrm{v}_{f_o|\varphi}\mathrm{C}_\varphi L_{\frac{\gamma}{4|\mathcal{K}|}}\frac{\sqrt{\mathbf{m}_{k^\circ}}}{n} \qquad (4.37)\\
\leq &\left(18\frac{L_{\gamma/8}}{\gamma}\mathrm{C}_\varphi \mathrm{v}_{f_o|\varphi}(p) + 138986\frac{\sqrt{2k^\circ}|\mathcal{K}|^{1/(p-1)}}{n\delta_{\mathcal{K}}^4}\frac{L^{5/2}_{\gamma/8}}{\gamma}\right)\frac{\sqrt{\mathbf{v}_{k^\circ}}}{\delta_{\mathcal{K}} n} \\
&+ 52\mathrm{v}_{f_o|\varphi}\mathrm{C}_\varphi \frac{L_{\gamma/8}}{\gamma}\frac{\sqrt{\mathbf{m}_{k^\circ}}}{n\delta_{\mathcal{K}}^2} \\
\leq &\left(70\frac{L_{\gamma/8}}{\gamma}\mathrm{C}_\varphi \mathrm{v}_{f_o|\varphi}(p) + 138986\frac{\sqrt{2k^\circ}|\mathcal{K}|^{1/(p-1)}}{n\delta_{\mathcal{K}}^4}\frac{L^{5/2}_{\gamma/8}}{\gamma}\right)\frac{\delta_{\mathcal{K}}^{-1}\sqrt{\mathbf{v}_{k^\circ}} \vee \delta_{\mathcal{K}}^{-2}\sqrt{\mathbf{m}_{k^\circ}}}{n}. \qquad (4.38)
\end{aligned}$$

Inserting in Equation (4.36) the definition of A_γ and using that $\frac{\sqrt{2k^\circ}|\mathcal{K}|^{1/(p-1)}}{\delta_{\mathcal{K}}^4 n} \leq \mathrm{d}\,\eta^2_{k^\circ}$, $\mathrm{r}_{\mathcal{K},\mathrm{s}} = r_{k^\circ,\mathrm{s}}$, $\frac{|\mathcal{K}|^{1/(p-1)}}{\delta_{\mathcal{K}}^4} \geq 1$ and $\delta_{\mathcal{K}} \in (0,1)$ implies

$$\begin{aligned}
q^2_{k^\circ}(f - f_o) \geq &\left(R^2 + 140\frac{L_{\gamma/8}}{\gamma}\mathrm{C}_\varphi \mathrm{v}_{f_o|\varphi}(p) + 260\frac{L_{\gamma/8}}{\gamma}\mathrm{C}_\varphi\,\mathrm{v} + 833934\frac{L^{5/2}_{\gamma/8}}{\gamma}\mathrm{d}\,\eta^2_{k^\circ}\right) \\
&\cdot\left(\frac{\mathrm{w}^2(k^\circ)}{\mathrm{s}^2(k^\circ)} \vee \frac{1}{n}\left(\delta_{\mathcal{K}}^{-1}\sqrt{\mathbf{v}_{k^\circ}} \vee \delta_{\mathcal{K}}^{-2}\sqrt{\mathbf{m}_{k^\circ}}\right)\right) - \frac{\mathrm{w}^2(k^\circ)}{\mathrm{s}^2(k^\circ)}R^2 \\
\geq &\left((2\cdot 70)\frac{L_{\gamma/8}}{\gamma}\mathrm{C}_\varphi \mathrm{v}_{f_o|\varphi}(p) + (2\cdot 18 + 2\cdot 112)\frac{L_{\gamma/8}}{\gamma}\overline{\mathrm{v}}_{f|\varphi}\mathrm{C}_\varphi\right. \\
&\left. + (2+4)138986\frac{L^{5/2}_{\gamma/8}}{\gamma}\frac{\sqrt{2k^\circ}|\mathcal{K}|^{1/(p-1)}}{\delta_{\mathcal{K}}^4 n}\right)\cdot\frac{1}{n}\left(\delta_{\mathcal{K}}^{-1}\sqrt{\mathbf{v}_{k^\circ}} \vee \delta_{\mathcal{K}}^{-2}\sqrt{\mathbf{m}_{k^\circ}}\right).
\end{aligned}$$

Finally, using the lower bound of the quantile $\tau_{k^\circ|\gamma/2}$ derived in Equation (4.38) we obtain

$$\begin{aligned}
q^2_{k^\circ}(f - f_o) \geq\ & 2\tau_{k^\circ|\gamma/2} \\
&+ 2\left(18\mathrm{C}_\varphi\overline{\mathrm{v}}_{f|\varphi} + 2\cdot 138986\frac{L_{\gamma/8}}{\gamma}\frac{\sqrt{2k^\circ}}{n}\right)L^{1/2}_{\gamma/8}\frac{\sqrt{\mathbf{v}_{k^\circ}}}{n} \\
&+ 2\cdot 112\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\frac{L_{\gamma/8}}{\gamma}\frac{\sqrt{\mathbf{m}_{k^\circ}}}{n}.
\end{aligned}$$

This shows Equation (4.35) and completes the proof. □

Proposition 4.4.3 yields $\mathrm{r}_{\mathcal{K},\mathrm{s}}(n)$, defined in Equation (4.33), as an upper bound for the radius of testing whenever $\eta^2_{\mathcal{K}|n}$ is equal to one for each $\mathrm{s} \in \mathcal{S}$. This is equivalent to

$$\sqrt{2k_{\mathcal{K},\mathrm{s}}(n)}|\mathcal{K}|^{1/(p-1)} \leq \mathrm{d}\, n\delta^4_{\mathcal{K}}.$$

Before establishing in which cases this inequality is fulfilled, we first discuss the rate $\mathrm{r}_{\mathcal{K},\mathrm{s}}(n)$ and the impact of its deviations from $\rho^2_{*,\mathrm{s}}(n)$ as well as choices for the collection $\mathcal{K}$.

Minimization over $\mathcal{K}$ For $\mathrm{r}_{\mathcal{K},\mathrm{s}}(n)$, we minimize over the collection $\mathcal{K}$ instead of $\mathbb{N}$. Ideally, $\mathcal{K} \subseteq \mathbb{N}$ is chosen such that its elements approximate the optimal parameter k^*_s given in Equation (4.27) sufficiently well for all $\mathrm{s} \in \mathcal{S}$. For the examples considered in Section 4.3, we have that $k^*_\mathrm{s} \leq n^2$ for $n \in \mathbb{N}$ large enough, refer to Table 4.1. Hence, a naive choice is $\mathcal{K} = [\![n^2]\!]$ with $|\mathcal{K}| = n^2$, which yields a factor $\delta^{-1}_{\mathcal{K}} = (1+\log|\mathcal{K}|)^{1/2}$ of order $(\log n)^{1/2}$. Denote here $\tilde{k} := k_{\mathcal{K},\mathrm{s}}(n)$ and

$$\tilde{\mathrm{r}}^2_{\mathcal{K},\mathrm{s}}(n) := \frac{1}{n}\left(\delta^{-1}_{\mathcal{K}}\sqrt{\mathbf{v}_{\tilde{k}}} \vee \delta^{-2}_{\mathcal{K}}\sqrt{\mathbf{m}_{\tilde{k}}}\right).$$

The resulting rates for the examples presented inSection 4.3 are given in Table 4.2. Further assume from now on that for the (o.s.-o.s.) case (first line) that $\sigma + a > -\frac{1}{4}$. Note that in the example of both densities to be ordinarily smooth (first line) and the case of super smoothness of s and ordinary smoothness of the error density (third line), the rate worsens by a logarithmic factor, while the case of an ordinary smooth regularity s and super smoothness of the error density (second line), the rate stays the same as in the non-adaptive case, see Table 4.1. Consequently, we will not consider this case any further in this section. In the fourth column $\kappa \in \mathbb{R}_+$ is some constant depending on a and σ. However, the exponential term dominates.

$\mathrm{s}(t)$	$\lvert\mathcal{M}_c[\varphi](t)\rvert$	$\frac{\mathrm{w}^2(k)}{\mathrm{s}^2(k)}$	$\tilde{\mathrm{r}}^2_{\mathcal{K},\mathrm{s}}(n)$	$k_{\mathcal{K},\mathrm{s}}(n)$	$\mathrm{r}^2_{\mathcal{K},\mathrm{s}}(n)$
$(1+t^2)^{\frac{s}{2}}$	$(1+t^2)^{-\frac{\sigma}{2}}$	$k^{-2(s-a)}$	$\frac{\log n}{n}\tilde{k}^{2(\sigma+a)+\frac{1}{2}}$	$\left(\frac{n}{\log n}\right)^{\frac{2}{4s+4\sigma+1}}$	$\left(\frac{n}{\log n}\right)^{\frac{-4(s-a)}{4s+4\sigma+1}}$
$(1+t^2)^{\frac{s}{2}}$	$\exp(-\lvert t\rvert^\sigma)$	$k^{-2(s-a)}$	$\frac{n}{\log n}\tilde{k}^{\kappa}\exp(2\tilde{k}^\sigma)$	$(\log n)^{\frac{1}{\sigma}}$	$(\log n)^{-\frac{2(s-a)}{\sigma}}$
$\exp(\lvert t\rvert^s)$	$(1+t^2)^{-\frac{\sigma}{2}}$	$\exp(2k^s)$	$\frac{n}{\log n}\tilde{k}^{2(\sigma+a)+\frac{1}{2}}$	$(\log n)^{\frac{1}{s}}$	$\frac{\log(n)}{n}(\log n)^{\frac{2(\sigma+a)+1/2}{s}}$

Table 4.2: Order of the radius of testing for $\mathrm{w}(t) \sim t^a$ for $t \in \mathbb{R}$, $a \in \mathbb{R}$ and $\mathcal{K} = [\![n^2]\!]$.

Discussion of adaptive factor Evidently, since $\delta_{\mathcal{K}} \in (0,1)$ we have

$$\rho^2_{\mathcal{K},s}(\delta_{\mathcal{K}}n) \leq \mathrm{r}^2_{\mathcal{K},\mathrm{s}}(n) \leq \rho^2_{\mathcal{K},s}(\delta^2_{\mathcal{K}}n).$$

Provided that the collection $\mathcal{K}$ is chosen such that $\rho^2_{\mathcal{K},s}(\delta_{\mathcal{K}}n) \asymp \rho^2_{\mathbb{N},s}(\delta_{\mathcal{K}})$ and $\rho^2_{\mathcal{K},s}(\delta^2_{\mathcal{K}}n) \asymp \rho^2_{\mathbb{N},s}(\delta_{\mathcal{K}}n)$ in the best case the effective sample size is $\delta_{\mathcal{K}}n$ and in the worst case $\delta^2_{\mathcal{K}}n$. As discussed above, in the case of $\mathcal{K} = [\![n^2]\!]$, the optimal parameter is included and the minimization is the same as over the natural numbers $\mathbb{N}$.

Choice of collection $\mathcal{K}$ In most cases, a finer grid than $\mathcal{K} = [\![n^2]\!]$ approximates the minimization over $\mathbb{N}$ well enough. More precisely, we consider a minimization over a geometric grid

$$\mathcal{K}_g := \{1\} \cup \{2^j, j \in [\![\log(n^2)]\!]\}.$$

Choosing $\mathcal{K} = \mathcal{K}_g$ results in an adaptive factor $\delta_{\mathcal{K}}^{-1}$ of oder $(\log \log n)^{1/2}$. Denote

$$\mathrm{r}^2_{\mathcal{K},1}(x) := \arg\min_{k \in \mathcal{K}} \left(\frac{\mathrm{w}^2(k)}{\mathrm{s}^2(k)} \vee \frac{1}{x} \delta_{\mathcal{K}}^{-1} \sqrt{\mathbf{v}_k} \right),$$
$$\mathrm{r}^2_{\mathcal{K},2}(x) := \arg\min_{k \in \mathcal{K}} \left(\frac{\mathrm{w}^2(k)}{\mathrm{s}^2(k)} \vee \frac{1}{x} \delta_{\mathcal{K}}^{-2} \sqrt{\mathbf{m}_k} \right).$$

Then, it holds that $\mathrm{r}^2_{\mathcal{K},\mathrm{s}}(n) = \mathrm{r}^2_{\mathcal{K},1}(n) \vee \mathrm{r}^2_{\mathcal{K},2}(n)$, by Balancing Lemma (Lemma A.2.1) in Schluttenhofer (2020), since $\frac{\mathrm{w}^2(k)}{\mathrm{s}^2(k)}$ is by Assumption 4.1.2 monotonically non-increasing $\frac{1}{n}\delta_{\mathcal{K}}^{-1}\sqrt{\mathbf{v}_k}$ and $\frac{1}{n}\delta_{\mathcal{K}}^{-2}\sqrt{\mathbf{m}_k}$ are monotonically non-decreasing in k. Table 4.3 gives for $\mathcal{K}_g$ the order of both $\mathrm{r}^2_{\mathcal{K},1}(n)$ and $\mathrm{r}^2_{\mathcal{K},2}(n)$, for the case that the Mellin transform of the error density is ordinarily smooth and the density f ordinarily smooth (first line) or super smooth (second line).

$\mathrm{s}(t)$	$\lvert\mathcal{M}_c[\varphi](t)\rvert$	$\mathrm{r}^2_{\mathcal{K},1}(n)$	$\mathrm{r}^2_{\mathcal{K},2}(n)$
$(1+t^2)^{-\frac{s}{2}}$	$(1+t^2)^{-\frac{\sigma}{2}}$	$\left(\frac{n}{(\log\log n)^{1/2}}\right)^{-\frac{4(s-a)}{4s+4\sigma+1}}$	$\left(\frac{n}{\log\log n}\right)^{-\frac{4(s-a)}{4s+4\sigma}}$
$\exp(\lvert t\rvert^s)$	$(1+t^2)^{-\frac{\sigma}{2}}$	$\frac{(\log\log n)^{1/2}}{n}(\log n)^{\frac{2(\sigma+a)+1/2}{s}}$	$\frac{\log\log n}{n}(\log n)^{\frac{2(\sigma+a)}{s}}$

Table 4.3: Order of the radius of testing $\mathrm{w}(t) \sim t^a$ for $t \in \mathbb{R}$, $a \in \mathbb{R}$ and $\mathcal{K} = \mathcal{K}_g$.

Note, that $\mathrm{r}^2_{\mathcal{K},2}(n)$ is asymptotically negligible compared with $\mathrm{r}^2_{\mathcal{K},1}(n)$. Hence, the upper bound in Proposition 4.4.3 asymptotically reduces to $\mathrm{r}^2_{\mathcal{K},1}(n)$ for both the first and the third line.

In a mildly ill-posed model with super smoothness, the smaller geometric collection

$$\mathcal{K}_m = \{1\} \cup \{2^j, j \in [\![m^{-1} \log\log n]\!]\},$$

for $m > 0$ is still sufficient. Then, the adaptive factor becomes of order $(\log\log\log n)^{1/2}$. The corresponding rates are given in Table 4.4. Again, the term $\mathrm{r}^2_{\mathcal{K},2}(n)$ is asymptotically negligible compared with $\mathrm{r}^2_{\mathcal{K},1}(n)$ and the upper bound in Proposition 4.4.3 asymptotically reduces to $\mathrm{r}^2_{\mathcal{K},1}(n)$. Note that the remark holds only for large n so that the adaptive factor is larger than one.

$\mathrm{s}(t)$	$\lvert\mathcal{M}_c[\varphi](t)\rvert$	$\mathrm{r}^2_{\mathcal{K},1}(n)$	${}^2_{\mathcal{K},2}(n)$
$\exp(-\lvert t\rvert^s)$	$(1+t^2)^{-\frac{\sigma}{2}}$	$\frac{(\log\log\log n)^{1/2}}{n}(\log n)^{\frac{2(\sigma+a)+1/2}{s}}$	$\frac{\log\log\log n}{n}(\log n)^{\frac{2(\sigma+a)}{s}}$

Table 4.4: Order of the radius of testing $\mathrm{w}(t) \sim t^a$ for $t \in \mathbb{R}$, $a \in \mathbb{R}$ and $\mathcal{K} = \mathcal{K}_m$.

For more details on the derivation of the rates in Tables 4.3 and 4.4, we refer to the calculations of Illustration 4.3.6. Schluttenhofer (2020), where analogous rates are calculated for the circular convolution model.

Order of factor $\eta^2_{\mathcal{K}|n}$ Let us discuss under which assumptions the factor $\eta^2_{\mathcal{K}|n}$ is of order one for the given examples such that $\mathrm{r}^2_{\mathcal{K},\mathrm{s}}(n)$ describes indeed the order of the radius of testing. First note that in all discussed collections it holds $|\mathcal{K}| \leq n^2$ and thus $\delta_{\mathcal{K}}^{-4} \leq (\log n)^2$. Let us start with the ordinary smooth - mildly ill-posed case (second line of Table 4.2). In this case, it holds that

$$\frac{\sqrt{2k_{\mathcal{K},\mathrm{s}}(n)}|\mathcal{K}|^{1/(p-1)}}{\mathrm{d}\, n\delta_{\mathcal{K}}^4} \lesssim \frac{n^{\frac{1}{4s+4\sigma+1}} n^{\frac{2}{(p-1)}}}{n(\log n)^{-2}}.$$

This is of order less or equal to one whenever

$$\frac{2}{p-1} - 1 + \frac{1}{4s+4\sigma+1} < 0.$$

For example, if $\frac{1}{4} < \sigma + s$ and $p \geq 5$, this is satisfied. Consequently, if the regularity parameters are sufficiently large and the moment condition Assumption 4.4.1 is satisfied for φ and f_o, then the parameter $\eta^2_{\mathcal{K}|n}$ is of order one. For the other two examples, line two and three of Table 4.2, we have that $\sqrt{k_{\mathcal{K},\mathrm{s}}(n)}$ is of logarithmic order, and, consequently, $p \geq 5$ is sufficient for $\eta^2_{\mathcal{K}|n} \sim 1$ to hold, as well. That is, the radius of testing is of order as given in Tables 4.2 to 4.4, depending on the choice of $\mathcal{K}$.

4.5 Simulation study

In this section, we illustrate the behavior of the test $\Delta_{k,\alpha}$ presented in Equation (4.21) and the data-driven max-test $\Delta_{\mathcal{K}|\gamma/2}$ defined in Equation (4.31).

To do so we consider the following example appearing also in the simulation study of the quadratic functional estimation, see Section 3.5. That is, we set $c = 0.5$ and $\mathrm{w}(t) = 1$ for $t \in \mathbb{R}$. Further, we consider as null hypothesis the case of an super-smooth null hypothesis density and ordinarily smooth error density φ. Specifically, we set the null hypothesis to be log-normally distributed with parameters $\mu = 0, \sigma^2 = 1$, i.e., for $x \in \mathbb{R}_+$

$$f_o(x) = \frac{1}{\sqrt{2\pi x^2}} \exp(-\log(x)^2/2)$$

and U being Pareto-distributed, i.e.

$$\varphi(x) = \mathbb{1}_{(1,\infty)}(x)x^{-2}.$$

For the random variable X we look both at the case that the null hypothesis is satisfied, i.e., $f = f_o$, and, at the case where the density of X belongs to the alternative. More precisely, we consider for X the two cases

$$f_1(x) = f_o(x),$$
$$f_2(x) = 2x\mathbb{1}_{(0,1)}(x).$$

The corresponding Mellin-transforms are given by

$$\mathcal{M}_{\frac{1}{2}}[f_1](t) = 2/(1.5 + 2\pi i t),$$
$$\mathcal{M}_{\frac{1}{2}}[f_2](t) = \exp\left((-0.5 + 2\pi i t)^2/2\right),$$
$$\mathcal{M}_{\frac{1}{2}}[\varphi](t) = (1.5 - 2\pi i t)^{-1}.$$

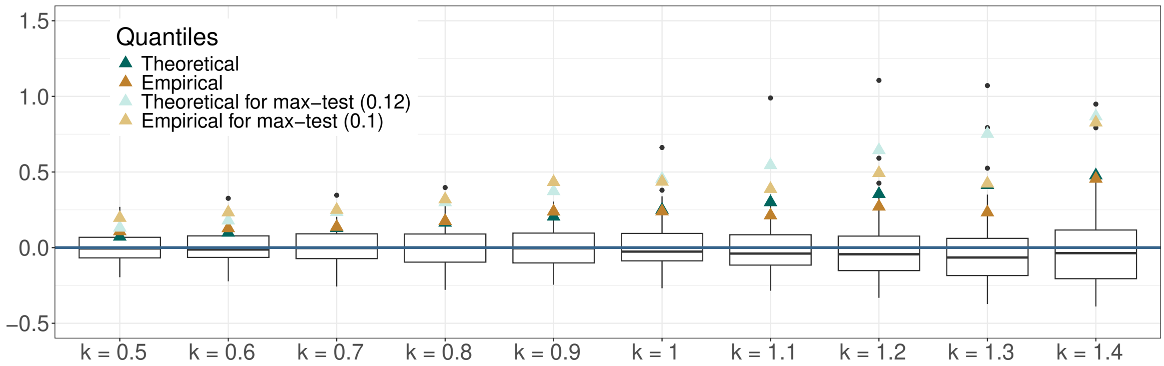
We refer again to Brenner Miguel (2023) for more details on Mellin transforms. The separation of the density to the null hypothesis is obviously zero for f_1 , that is $q^2(f_1 - f_o) = 0$. For the second density f_2 we have $q^2(f_2 - f_o) = 0.5$.

Below, we show the results of our simulation study in Figure 4.1 and Figure 4.2. In Figure 4.1 for the first example, $f = f_1$, and in Figure 4.2 for second example, $f = f_2$, the boxplots of the values of the test statistic $\hat{q}_k^2$ are depicted over 50 iterations for $k \in \{0.5, 0.6, \dots, 1.4\}$ and $n \in \{100, 500\}$. The examples were chosen such that the densities and their Mellin transforms have a simple representation. However, the bias is for $k \leq 1.4$ already too small to see variation in the estimates. Minimizing over a subset of $\mathbb{N}$, we would always choose $k = 1$ for the given sample sizes. Only immensely higher sample sizes would lead to a different output. Thus, we rescaled the set of dimension parameters on the interval $[0.5, 1.4]$.

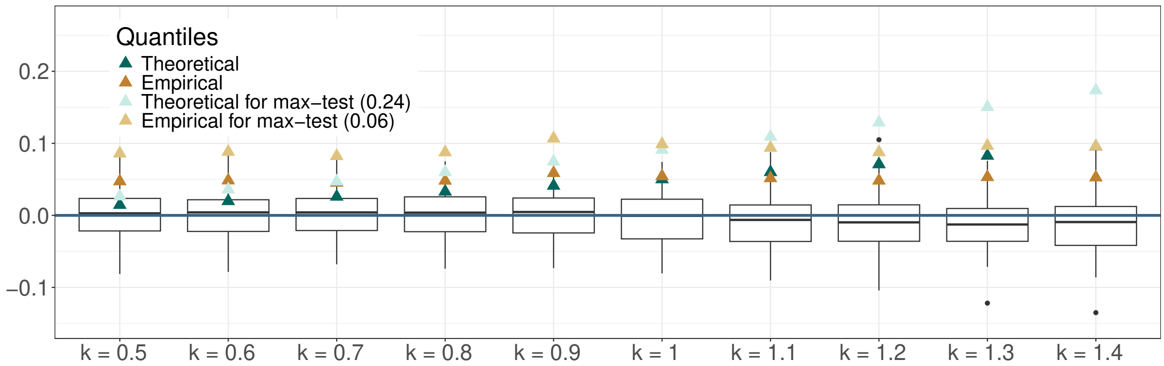
In both Figures 4.1 and 4.2, the triangles indicate different estimates for α-quantiles for $\alpha = 0.1$. The dark blue triangle indicates the quantile $\tau_k(0.1)$ proposed in Equation (4.21) and the light blue one the corresponding quantile $\tau_{k|0.1}$ for the max-test defined in Equation (4.30) for $\mathcal{K} = \{0.5, 0.6, \dots, 1.4\}$. Note that since $L_{\gamma/(4|\mathcal{K}|)} \leq L_{\gamma/4}\delta_{\mathcal{K}}^{-2}$ (see proof of Proposition 4.4.3), theses theoretical quantiles differ by $\delta_{\mathcal{K}}^{-1} = (1+\log(10))^{1/2} \approx 1.82$ up to constants. In dark brown, the empirical 0.1-quantiles under the null hypothesis are indicated. The light brown triangles give the empirical quantiles multiplied by the factor $\delta_{\mathcal{K}}^{-1}$ to give an empirical quantile for the max-test. The numbers in brackets in the legend give the rejection rate of the corresponding max-tests for each example.

As already mentioned, $n = \{100, 500\}$ are relatively small sample sizes and the bias is relatively small compared to the variance. Consequently, we decide to use a value equal to 0.6 for the constants of both $\tau_k(0.1)$ and $\tau_{k|0.1}$. The effect of this can be seen in the boundaries of the interval considered for parameter k: Under the null hypothesis in Figure 4.1 we note that the test does not adhere to the significance level of 0.1 at the boundaries. Choosing the theoretical constants instead, the test would always reject for these examples.

However, close to $k = 1$ the theoretical quantiles appear to match the empirical quantiles. Looking at the alternative in Figure 4.2, the separation appears to be large enough such that the test has high power already for a sample size of $n = 500$.

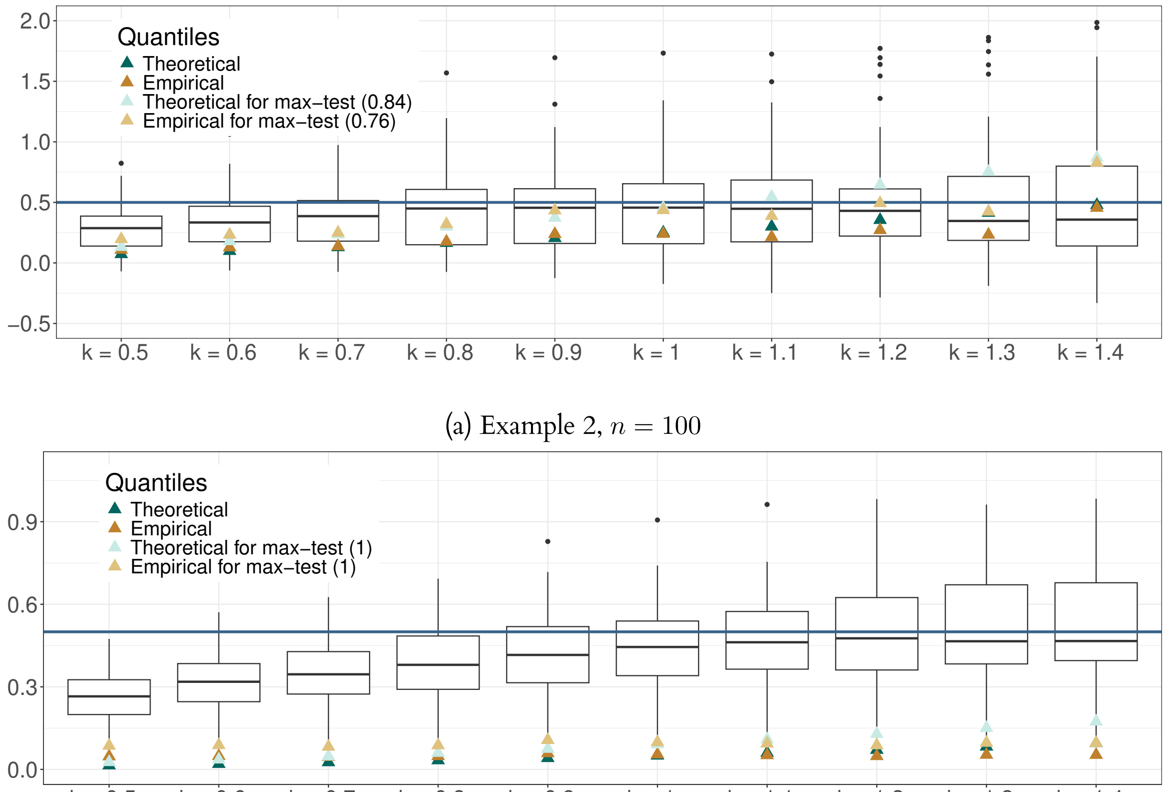


(a) Example 1, $n = 100$



(b) Example 1, $n = 500$

Figure 4.1: The boxplots represent the values of $\hat{q}_k^2$ for $f = f_1$ over 50 iterations. The horizontal lines indicate $q^2 = 0$. The triangles indicate different estimates for 0.1-quantiles. The numbers given in the legend indicate the rejection rate of the corresponding test.



(a) Example 2, $n = 100$

(b) Example 2, $n = 500$

Figure 4.2: The boxplots represent the values of $\hat{q}_k^2$ for $f = f_2$ over 50 iterations. The horizontal lines indicate $q^2 = 0.5$. The triangles indicate different estimates for 0.1-quantiles. The numbers given in the legend indicate the rejection rate of the corresponding test.

4.6 Auxiliary results

The next assertion, a concentration inequality for canonical U-statistics, provides our key argument in order to control the deviation of the test statistics. It is a reformulation of Theorem 3.4.8 in Giné and Nickl (2016). The result was used in the proofs of Chapter 3.

For $(X_j)_{j\in[\![n]\!]}$, $n \in \mathbb{N}$ independent and identically distributed random variables in $(\mathbb{R}_+, \mathscr{B}_+)$ and bounded symmetric kernel $h\colon \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$, a U-statistic is a sum of the form

$$U_n = \frac{1}{n(n-1)} \sum_{\substack{j_1 \neq j_2 \\ j_1, j_2 \in [\![n]\!]}} h(X_{j_1}, X_{j_2}).$$

The U-statistic is called canonical if for all $i, j \in [\![n]\!]$ and all $x, y \in \mathbb{R}_+$

$$\mathbb{E}\left[h(X_i, y)\right] = \mathbb{E}\left[h(x, X_j)\right] = 0.$$

In this chapter, use the following adaption of Theorem 3, Chap.1, from Lee (1990).

Lemma 4.6.1:
Let $U := \frac{1}{n(n-1)} \sum_{\substack{j \neq l \\ j,l \in [\![n]\!]}} h(X_j, X_l)$ be a U-statistic with i.i.d. random variables $(X_j)_{j\in[\![n]\!]}$. Then

$$\mathrm{Var}(U) \leq \frac{2}{n(n-1)} \mathbb{E}^2[|h(X_1, X_2)|^2].$$

For the following lemma we have used the notation $\|\cdot\|_{\mathbb{L}^\infty_{+,2}}$ to indicate the essential supremum for functions on $\mathbb{R}_+ \times \mathbb{R}_+$.

Proposition 4.6.2:
Let U_n be a canonical U-statistic for $(X_j)_{j\in[\![n]\!]}$, $n \geq 2$ i.i.d. $\mathbb{R}_+$-valued random variables and kernel $h\colon \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ bounded and symmetric. Let

$$\begin{aligned}
&A \geq \|h\|_{\mathbb{L}^\infty_{+,2}} \\
&B^2 \geq \|\mathbb{E}[h^2(X_1, \cdot)]\|_{\mathbb{L}^\infty_{+}} \\
&C^2 \geq \mathbb{E}[h^2(X_1, X_2)] \\
&D \geq \sup\{\mathbb{E}[h(X_1, X_2)\xi(X_1)\zeta(X_2)] : \mathbb{E}[\xi^2(X_1)] \leq 1, \mathbb{E}[\zeta^2(X_2)] \leq 1\}.
\end{aligned}$$

Then, for all $x \geq 0$

$$\mathbb{P}\left(U_n \geq 8\frac{C}{n}x^{1/2} + 13\frac{D}{n}x + 261\frac{B}{n^{3/2}}x^{\frac{3}{2}} + 343\frac{A}{n^2}x^2\right) \leq \exp(1-x).$$

4.7 Proofs for Section 4.2

The decomposition is similar to the estimation of the quadratic functional in Chapter 3.

4.7.1 U-statistic results

We apply the concentration inequality for canonical U-statistics given in Proposition 4.6.2. First, we see that U_k given in Equation (4.14) can be written as

$$U_k = \frac{1}{n(n-1)} \sum_{\substack{j_1 \neq j_2 \\ j_1, j_2 \in [\![n]\!]}} h_k(Y_{j_1}, Y_{j_2}) \tag{4.39}$$

with $h_k(x,y) := \int_{\mathbb{R}} \mathbb{1}_{[-k,k]} |\mathcal{M}_c[\varphi]|^{-2} \Psi(x,\cdot)\overline{\Psi}(y,\cdot)\, \mathrm{w}^2\, d\lambda$ for $x, y \in \mathbb{R}_+$. Note that for $x \in \mathbb{R}_+$ the term $|x^{c-1+2\pi it}| = x^{c-1}$ is not bounded and hence $|h_k(x,y)|$ is generally not bounded for $x, y \in \mathbb{R}_+$. Therefore, we decompose h_k in a bounded and a remaining unbounded part. More precisely, given $\delta \in \mathbb{R}_+$ specified later, we denote for $y \in \mathbb{R}_+$ and $t \in \mathbb{R}$

$$\psi_k^b(y,t) := \mathbb{1}_{[0,\delta]}(y^{c-1}) y^{c-1+2\pi it} \quad \text{and} \quad \psi_k^u(y,t) := \mathbb{1}_{(\delta,\infty)}(y^{c-1}) y^{c-1+2\pi it}.$$

Define the bounded part of kernel h_k as

$$h_k^b(x,y) := \int_{-k}^{k} \frac{(\psi_k^b(y,t) - \mathbb{E}_g[\psi_k^b(Y_1,t)])(\psi_k^b(x,-t) - \mathbb{E}_g[\psi_k^b(Y_1,-t)])}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t). \tag{4.40}$$

Then, h_k^b is indeed bounded since $|\psi_k^b(y,t)| \leq \delta$. Analogously, define

$$\begin{aligned} h_k^u(x,y) := & \int_{-k}^{k} \frac{(\psi_k^u(y,t) - \mathbb{E}_g[\psi_k^u(Y_1,t)])(\psi_k^u(x,-t) - \mathbb{E}_g[\psi_k^u(Y_1,-t)])}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t) \\ & + \int_{-k}^{k} \frac{(\psi_k^u(y,t) - \mathbb{E}_g[\psi_k^u(Y_1,t)])(\psi_k^b(x,-t) - \mathbb{E}_g[\psi_k^b(Y_1,-t)])}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t) \\ & + \int_{-k}^{k} \frac{(\psi_k^b(y,t) - \mathbb{E}_g[\psi_k^b(Y_1,t)])(\psi_k^u(x,-t) - \mathbb{E}_g[\psi_k^u(Y_1,-t)])}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t). \end{aligned} \tag{4.41}$$

Then, h_k^b and h_k^u are also symmetric and real-valued and we have $h_k = h_k^b + h_k^u$. Denote by U_k^b and U_k^u the corresponding canonical U-statistics, respectively. Then, it holds $U_k = U_k^b + U_k^u$. We first derive a quantile for the bounded part U_k^b in Lemma 4.7.2 using Proposition 4.6.2 and, subsequently, a quantile for the unbounded part U_k^u in Lemma 4.7.3. Both results are combined to derive the quantile of U_k in Lemma 4.7.4. In this section, for $\varphi, h \in \mathbb{L}^1_+(\mathrm{x}^{2p(c-1)})$ we use the following notation

$$\mathrm{v}_{h|\varphi}(p) = (\|h \circledast \varphi\|_{\mathbb{L}^1_+(\mathrm{x}^{2p(c-1)})} \vee 1).$$

For the bounded U-statistic, the following lemma gives the constants for applying Proposition 4.6.2, which was shown in Chapter 3.

Lemma 4.7.1 (Constants for the bounded U-statistic):
Let the assumptions of Lemma 4.7.4 be satisfied. The kernel h_k^b introduced in Equation (4.40) with $\delta \in \mathbb{R}_+$ is real-valued, bounded, symmetric and satisfies for all $y \in \mathbb{R}_+$ that $\mathbb{E}_g[h_k^b(Y, y)] = 0$. Then

$$\begin{aligned} A &:= 4\delta^2 \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^2)}, \\ B^2 &:= 4\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} \delta^2 \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^4(\mathrm{w}^4)}, \\ C^2 &:= \mathrm{C}_\varphi \mathrm{v}^2_{f|\varphi} \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^4(\mathrm{w}^4)}, \\ D &:= 4\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} \|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^\infty(\mathrm{w})} \end{aligned}$$

satisfy the conditions of the U-statistic concentration inequality given in Proposition 4.6.2.

With this result we show the following lemma for the bounded part of the U-statistic Equation (4.14).

Lemma 4.7.2 (Quantile of the bounded U-statistic):
Let the assumptions of Lemma 4.7.4 be satisfied. Consider for each $n \geq 2$ and any $\delta \in \mathbb{R}_+$ the canonical U-statistic U_k^b. For $\gamma \in (0, 1)$ and

$$\begin{aligned} \tau_k^b(\gamma) := &\left(9\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} + 69493\frac{\delta^2}{n}\sqrt{2k}L_\gamma^2\right) L_\gamma^{1/2} \frac{\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^4(\mathrm{w}^4)}}{n} \\ &+ 52\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} L_\gamma \frac{\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^\infty(\mathrm{w})}}{n} \end{aligned}$$

we have

$$\mathbb{P}^n_{\varphi,f}(U_k^b \geq \tau_k^b(\gamma)) \leq \gamma.$$

Proof of Lemma 4.7.2. We intend to apply the exponential inequality for canonical U-statistics given in Proposition 4.6.2 using quantities A, B, C, and D given in Proposition 4.6.2 which verify the four required inequalities. Precisely, for $\gamma \in (0, 1)$ and $x = L_\gamma = 1 - \log(\gamma)$ due to Proposition 4.6.2 we have

$$\mathbb{P}^n_{\varphi,f}\left(U_k \geq 8\frac{C}{n}L_\gamma^{1/2} + 13\frac{D}{n}L_\gamma + 261\frac{B}{n^{3/2}}L_\gamma^{3/2} + 343\frac{A}{n^2}L_\gamma^2\right) \leq \exp(1 - L_\gamma) = \gamma.$$

Exploiting Lemma 4.7.2 we have that

$$\begin{aligned} &8\frac{C}{n}L_\gamma^{1/2} + 13\frac{D}{n}L_\gamma + 261\frac{B}{n^{3/2}}L_\gamma^{3/2} + 343\frac{A}{n^2}L_\gamma^2 \\ &= 8\frac{\mathrm{C}_\varphi^{1/2}\mathrm{v}_{f|\varphi}\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^4(\mathrm{w}^4)}}{n}L_\gamma^{1/2} + 52\frac{\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^\infty(\mathrm{w})}}{n}L_\gamma \\ &\quad + 522\frac{\mathrm{C}_\varphi^{1/2}\mathrm{v}_{f|\varphi}^{1/2}\delta\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^4(\mathrm{w}^4)}}{n^{3/2}}L_\gamma^{3/2} + 1372\frac{\delta^2\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^2)}}{n^2}L_\gamma^2 \end{aligned}$$

which together with $\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^2(\mathrm{w}^2)} \leq (2k)\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^4_{\mathbb{L}^4(\mathrm{w}^4)}$ implies

$$\begin{aligned} &8\frac{C}{n}L_\gamma^{1/2} + 13\frac{D}{n}L_\gamma + 261\frac{B}{n^{3/2}}L_\gamma^{3/2} + 343\frac{A}{n^2}L_\gamma^2 \\ &\leq 52\frac{\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^\infty(\mathrm{w})}}{n}L_\gamma + \frac{\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^4(\mathrm{w}^4)}}{n}L_\gamma^{1/2} \\ &\quad \cdot \left(8\mathrm{C}_\varphi^{1/2}\mathrm{v}_{f|\varphi} + 522\frac{\mathrm{C}_\varphi^{1/2}\mathrm{v}_{f|\varphi}^{1/2}\delta}{n^{1/2}}L_\gamma + 1372\frac{\delta^2\sqrt{2k}}{n}L_\gamma^{3/2}\right). \end{aligned}$$

Using that $2ab \leq a^2 + b^2$ for $a, b \in \mathbb{R}_+$, we have that

$$522\frac{\mathrm{C}_\varphi^{1/2}\mathrm{v}_{f|\varphi}^{1/2}\delta}{n^{1/2}}L_\gamma \leq \mathrm{C}_\varphi \mathrm{v}_{f|\varphi} + (522/2)^2\frac{\delta^2 L_\gamma^2}{n}$$

and, thus, since $\sqrt{2k}, L_\gamma \geq 1$

$$\begin{aligned} &8\frac{C}{n}L_\gamma^{1/2} + 13\frac{D}{n}L_\gamma + 261\frac{B}{n^{3/2}}L_\gamma^{3/2} + 343\frac{A}{n^2}L_\gamma^2 \\ &\leq 52\frac{\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^\infty(\mathrm{w})}}{n}L_\gamma + \frac{\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^4(\mathrm{w}^4)}}{n}L_\gamma^{1/2} \\ &\quad \cdot \left(9\mathrm{C}_\varphi \mathrm{v}_{f|\varphi} + ((522/2)^2 + 1372)\frac{\delta^2\sqrt{2k}}{n}L_\gamma^2\right) = \tau_k^b(\gamma) \end{aligned}$$

with $(522/22)^2 + 1372 = 69493$ due to Proposition 4.6.2 follows

$$\mathbb{P}^n_{\varphi,f}(U_k^b \geq \tau_k^b(\gamma)) \leq \gamma$$

which completes the proof. □

Further, we give a quantile of the unbounded part of U-statistic Equation (4.14) of order $1/n$.

Lemma 4.7.3 (Quantile of the unbounded U-statistic)**:**
Let the assumptions of Lemma 4.7.4 be satisfied. Consider for each $n \geq 2$ and $\delta \in \mathbb{R}_+$ the canonical U-statistic U_k^u. For $\gamma \in (0, 1)$ setting

$$\tau_k^u(\gamma) := \frac{\delta^{(1-p)}}{\gamma^{1/2}}6\mathrm{C}_\varphi^{1/2}\mathrm{v}_{f|\varphi}^{1/2}(\mathrm{v}_{f|\varphi}(p))^{1/2}\frac{\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^4(\mathrm{w}^4)}}{n}$$

we have

$$\mathbb{P}^n_{\varphi,f}(U_k^u \geq \tau_k^u(\gamma)) \leq \gamma.$$

Proof of Lemma 4.7.3. We start the proof with the observation that the symmetric and real-valued kernel h_k^u given in Equation (4.41) satisfies

$$\begin{aligned}
&\mathbb{E}_g^2[|h_k^u(Y_1,Y_2)|^2] \\
&\leq 3\mathbb{E}_g^2\left[\left|\int_{-k}^{k}\frac{(\psi_k^u(Y_2,t)-\mathbb{E}_g[\psi_k^u(Y,t)])(\psi_k^u(Y_1,-t)-\mathbb{E}_g[\psi_k^u(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right] \\
&\quad+3\mathbb{E}_g^2\left[\left|\int_{-k}^{k}\frac{(\psi_k^u(Y_2,t)-\mathbb{E}_g[\psi_k^u(Y,t)])(\psi_k^b(Y_2,-t)-\mathbb{E}_g[\psi_k^b(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right] \\
&\quad+3\mathbb{E}_g^2\left[\left|\int_{-k}^{k}\frac{(\psi_k^b(Y_2,t)-\mathbb{E}_g[\psi_k^b(Y,t)])(\psi_k^u(Y_1,-t)-\mathbb{E}_g[\psi_k^u(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right]
\end{aligned}$$

Note that the second and the third summand are the same due to symmetry of the integrals. Thus, we obtain

$$\begin{aligned}
&\mathbb{E}_g^2[|h_k^u(Y_1,Y_2)|^2] \\
&= 3\mathbb{E}_g^2\left[\left|\int_{-k}^{k}\frac{(\psi_k^u(Y_2,t)-\mathbb{E}_g[\psi_k^u(Y,t)])(\psi_k^u(Y_1,-t)-\mathbb{E}_g[\psi_k^u(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right] \\
&\quad+6\mathbb{E}_g^2\left[\left|\int_{-k}^{k}\frac{(\psi_k^u(Y_2,t)-\mathbb{E}_g[\psi_k^u(Y,t)])(\psi_k^b(Y_1,-t)-\mathbb{E}_g[\psi_k^b(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right]. \qquad (4.42)
\end{aligned}$$

Consider the first summand. Due to the independence of Y_1 and Y_2 we first consider the expectation with respect to Y_2 for fixed $Y_1 = y \in \mathbb{R}_+$. We have that

$$\begin{aligned}
&\mathbb{E}_g\left[\left|\int_{-k}^{k}\frac{(\psi_k^u(Y_2,t)-\mathbb{E}_g[\psi_k^u(Y,t)])(\psi_k^u(y,-t)-\mathbb{E}_g[\psi_k^u(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right] \\
&\leq \mathbb{E}_g\left[\left|\int_{-k}^{k}\frac{\psi_k^u(Y_2,t)(\psi_k^u(y,-t)-\mathbb{E}_g[\psi_k^u(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right] \\
&= \mathbb{E}_g\left[\mathbb{1}_{(\delta,\infty)}(Y_2^{c-1})\left|\int_{-k}^{k}\frac{Y_2^{c-1+2\pi it}(\psi_k^u(y,-t)-\mathbb{E}_g[\psi_k^u(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right] \\
&\leq \mathbb{E}_g\left[\left|\int_{-k}^{k}\frac{Y_2^{c-1+2\pi it}(\psi_k^u(y,-t)-\mathbb{E}_g[\psi_k^u(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right].
\end{aligned}$$

Using Lemma 1.2.5 (i) with $h(t) := \frac{(\psi_k^u(y,-t)-\mathbb{E}_g[\psi_k^u(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)$, it follows for $y \in \mathbb{R}_+$

$$\begin{aligned}
&\mathbb{E}_g\left[\left|\int_{-k}^{k}\frac{Y_2^{c-1+2\pi it}(\psi_k^u(y,-t)-\mathbb{E}_g[\psi_k^u(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2}\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right] \\
&\leq \|g\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})}\|\mathbb{1}_{[-k,k]}\frac{(\psi_k^u(y,\cdot)-\mathbb{E}_g[\psi_k^u(Y,\cdot)])}{|\mathcal{M}_c[\varphi]|^2}\|^2_{\mathbb{L}^2(\mathrm{w}^2)}.
\end{aligned}$$

Similarly, for the second summand of Equation (4.42) we get for $y \in \mathbb{R}_+$ that

$$\begin{aligned}
&\mathbb{E}_g\left[\left|\int_{-k}^{k} \frac{(\psi_k^u(y,t) - \mathbb{E}_g[\psi_k^u(Y,t)])(\psi_k^b(Y_2,-t) - \mathbb{E}_g[\psi_k^b(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t)d\lambda(t)\right|^2\right] \\
&\leq \mathbb{E}_g^2\left[\left|\int_{-k}^{k} \frac{\psi_k^b(Y_2,t)(\psi_k^u(y,-t) - \mathbb{E}_g[\psi_k^u(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t)d\lambda(t)\right|^2\right] \\
&= \mathbb{E}_g\left[\mathbb{1}_{[0,\delta]}(Y_2^{c-1})\left|\int_{-k}^{k} \frac{Y_2^{c-1+2\pi it}(\psi_k^u(y,-t) - \mathbb{E}_g[\psi_k^u(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t)d\lambda(t)\right|^2\right] \\
&\leq \mathbb{E}_g\left[\left|\int_{-k}^{k} \frac{Y_2^{c-1+2\pi it}(\psi_k^u(y,-t) - \mathbb{E}_g[\psi_k^u(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t)d\lambda(t)\right|^2\right].
\end{aligned}$$

Using again Lemma 1.2.5 (i) with $h(t) := \frac{(\psi_k^u(y,-t) - \mathbb{E}_g[\psi_k^u(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t)$, it follows for $y \in \mathbb{R}_+$

$$\begin{aligned}
&\mathbb{E}_g\left[\left|\int_{-k}^{k} \frac{Y_2^{c-1+2\pi it}(\psi_k^u(y,-t) - \mathbb{E}_g[\psi_k^u(Y,-t)])}{|\mathcal{M}_c[\varphi](t)|^2} \mathrm{w}^2(t)d\lambda(t)\right|^2\right] \\
&\leq \|g\|_{\mathbb{L}_+^\infty(\mathrm{x}^{2c-1})}\|\mathbb{1}_{[-k,k]}\frac{(\psi_k^u(y,\cdot) - \mathbb{E}_g[\psi_k^u(Y,\cdot)])}{|\mathcal{M}_c[\varphi]|^2}\|_{\mathbb{L}^2(\mathrm{w}^4)}^2.
\end{aligned}$$

Note that the upper bounds for both summands of Equation (4.42) are the same. Further, we have that $\|g\|_{\mathbb{L}_+^\infty(\mathrm{x}^{2c-1})} \leq \mathrm{v}_{f|\varphi}\mathrm{C}_\varphi$ and for $p \geq 2$ it holds

$$\mathbb{E}_g[|\mathbb{1}_{(\delta,\infty)}(Y_1^{c-1})Y_1^{c-1+2\pi it}|^2] \leq \delta^{2(1-p)}\mathrm{v}_{f|\varphi}(p).$$

Combining this with the other derived bounds we get

$$\begin{aligned}
\mathbb{E}_g^2[|h_k^u(Y_1,Y_2)|^2] &\leq 9\mathbb{E}_g\left[\|g\|_{\mathbb{L}_+^\infty(\mathrm{x}^{2c-1})}\|\mathbb{1}_{[-k,k]}\frac{(\psi_k^u(Y_1,\cdot) - \mathbb{E}_g[\psi_k^u(Y,\cdot)])}{|\mathcal{M}_c[\varphi]|^2}\|_{\mathbb{L}^2(\mathrm{w}^4)}^2\right] \\
&\leq 9\mathrm{v}_{f|\varphi}\mathrm{C}_\varphi\int_{-k}^{k}\mathbb{E}_g\left[\left|\frac{(\psi_k^u(y,t) - \mathbb{E}_g[\psi_k^u(Y,t)])}{|\mathcal{M}_c[\varphi](t)|^2}\right|^2\right]\mathrm{w}^4(t)d\lambda(t) \\
&\leq 9\mathrm{v}_{f|\varphi}\mathrm{C}_\varphi\delta^{2(1-p)}\mathrm{v}_{f|\varphi}(p)\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^4(\mathrm{w}^4)}^4.
\end{aligned}$$

Applying Lemma 4.6.1 we immediately obtain

$$\begin{aligned}
\mathbb{E}_g^n\left[|U_k^u|^2\right] &= \frac{2}{n(n-1)}\mathbb{E}_g^2[|h_k^u(Y_1,Y_2)|^2] \leq \frac{36}{n^2}\mathrm{v}_{f|\varphi}\mathrm{C}_\varphi\delta^{2(1-p)}\mathrm{v}_{f|\varphi}(p)\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^4(\mathrm{w}^4)}^4 \\
&= \gamma|\tau_k^u(\gamma)|^2.
\end{aligned}$$

Hence, by applying Markov's inequality the result follows, i.e.,

$$\mathbb{P}_{\varphi,f}^n(U_k^u \geq \tau_k^u(\gamma)) \leq \frac{\mathbb{E}_g^n\left[|U_k^u|^2\right]}{|\tau_k^u(\gamma)|^2} \leq \gamma,$$

which completes the proof. □

Lemma 4.7.4 (Quantile of the U-statistic)**:**
Let Assumptions 4.1.1, 4.4.1 and 4.1.3 be satisfied. Consider for each $n \geq 2$ the canonical U-statistic U_k defined in Equation (4.39). For $p \geq 2$, $\gamma \in (0,1)$, $f \in \mathbb{L}^1_+(\mathrm{x}^{2p(c-1)})$ and

$$\tau^{U_k}(\gamma) := \left(18\mathrm{v}_{f|\varphi}(p)\mathrm{C}_\varphi + 69493\frac{\sqrt{2k}}{n\gamma^{1/(p-1)}}L_{\gamma/2}^{2-1/(p-1)}\right)\frac{\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^4(\mathrm{w}^4)}}{n}L_{\gamma/2}^{1/2} + 52\mathrm{C}_\varphi\mathrm{v}_{f|\varphi}\frac{\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^\infty(\mathrm{w})}}{n}L_{\gamma/2} \tag{4.43}$$

we have

$$\mathbb{P}^n_{\varphi,f}(U_k \geq \tau^{U_k}(\gamma)) \leq \gamma.$$

Proof of Lemma 4.7.4. We start the proof by decomposing with

$$\delta^2 = L_{\gamma/2}^{1/(1-p)}\gamma^{1/(1-p)} > 0 \tag{4.44}$$

the canonical U-statistic $U_k = U_k^b + U_k^u$ with kernels given in (4.40) and (4.41). We show below that $\tau^{U_k}(\gamma)$ given in (4.43) satisfies

$$\tau^{U_k}(\gamma) \geq \tau_k^b(\gamma/2) + \tau_k^u(\gamma/2). \tag{4.45}$$

Consequently, from Lemmata 4.7.2 and 4.7.3 we immediately obtain the result, that is,

$$\mathbb{P}^n_{\varphi,f}(U_k \geq \tau^{U_k}(\gamma)) \leq \mathbb{P}^n_{\varphi,f}(U_k^b \geq \tau_k^b(\gamma/2)) + \mathbb{P}^n_{\varphi,f}(U_k^u \geq \tau_k^u(\gamma/2)) \leq \gamma/2 + \gamma/2 = \gamma.$$

It remains to show (4.45). We have that

$$\begin{aligned}\tau_k^b(\gamma/2) + \tau_k^u(\gamma/2) = &\left(9\mathrm{C}_\varphi\mathrm{v}_{f|\varphi} + 69493\frac{\delta^2}{n}\sqrt{2k}L^2_{\gamma/2}\right)L_{\gamma/2}^{1/2}\frac{\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^4(\mathrm{w}^4)}}{n}\\ &+ 52\mathrm{C}_\varphi\mathrm{v}_{f|\varphi}L_{\gamma/2}\frac{\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^\infty(\mathrm{w})}}{n}\\ &+ \sqrt{2}\frac{\delta^{(1-p)}}{\gamma^{1/2}}6\mathrm{C}_\varphi^{1/2}\mathrm{v}_{f|\varphi}^{1/2}(\mathrm{v}_{f|\varphi}(p))^{1/2}\frac{\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^4(\mathrm{w}^4)}}{n}\end{aligned}$$

Using $\mathrm{v}_{f|\varphi}, \mathrm{v}^2_{f|\varphi} \leq \mathrm{v}_{f|\varphi}(p)$ yields

$$\begin{aligned}\tau_k^b(\gamma/2) + \tau_k^u(\gamma/2) \leq &\left(9\mathrm{C}_\varphi\mathrm{v}_{f|\varphi}(p) + 69493\frac{\delta^2}{n}\sqrt{2k}L^2_{\gamma/2} + L_{\gamma/2}^{-1/2}9\frac{\delta^{(1-p)}}{\gamma^{1/2}}\mathrm{C}_\varphi\mathrm{v}_{f|\varphi}(p)\right)\\ &\cdot L_{\gamma/2}^{1/2}\frac{\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^4(\mathrm{w}^4)}}{n}\\ &+ 52\mathrm{C}_\varphi\mathrm{v}_{f|\varphi}L_{\gamma/2}\frac{\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^\infty(\mathrm{w})}}{n}.\end{aligned}$$

With the choice of δ given in Equation (4.44), it holds $\delta^{1-p} = L_{\gamma/2}^{1/2}\gamma^{1/2}$ and it follows

$$\begin{aligned}\tau_k^b(\gamma/2)+\tau_k^u(\gamma/2) &\leq \left(18\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}(p)+69493\gamma^{-1/(p-1)}\frac{L_{\gamma/2}^{2-1/(p-1)}}{n}\sqrt{2k}\right)L_{\gamma/2}^{1/2}\frac{\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^4(\mathrm{w}^4)}}{n}\\ &\quad+52\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}L_{\gamma/2}\frac{\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^\infty(\mathrm{w})}}{n}\\ &= \tau^{U_k}(\gamma)\end{aligned}$$

which completes the proof. □

4.7.2 Linear statistic result

Consider the centered linear statistic W_k defined in Equation (4.15) admitting as kernel the real-valued function

$$h_k(y) := \int_{-k}^{k} |\mathcal{M}_c[\varphi](t)|^{-2}\Psi(y,t)(\overline{\mathcal{M}_c[g]}(t)-\overline{\mathcal{M}_c[g_o]}(t))\,\mathrm{w}^2(t)d\lambda(t).$$

Using Markov's inequality, we show the following result for its quantile.

Lemma 4.7.5 (Quantile of the linear statistic)**:**
Let Assumptions 4.1.1, 4.4.1 and 4.1.3 be satisfied. Further, assume that $f \in \mathbb{L}^1_+(\mathrm{x}^{2(c-1)})$. Consider for $n \in \mathbb{N}$ the linear statistic W_k defined in (4.15). For $\gamma \in (0,1)$ setting

$$\tau^{W_k}(\gamma) := \frac{1}{4}\|(\mathcal{M}_c[g]-\mathcal{M}_c[g_o])\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^2)}+\frac{\mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^\infty(\mathrm{w})}}{\gamma n}$$

we have

$$\mathbb{P}^n_{\varphi,f}(W_k \geq \tau^{W_k}(\gamma)) \leq \gamma.$$

Proof of Lemma 4.7.5. By making use of that fact that observations $(Y_j)_{j\in[\![n]\!]}$ are i.i.d. we get that

$$\begin{aligned}n\mathbb{E}^n_g[|W_k|^2] &= \mathbb{E}_g[|h_k(Y_1)|^2]\\ &= \mathbb{E}_g\left[\left|\int_{-k}^{k}\frac{1}{|\mathcal{M}_c[\varphi](t)|^2}\Psi(Y_1,t)(\overline{\mathcal{M}_c[g]}(t)-\overline{\mathcal{M}_c[g_o]}(t))\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right]\\ &\leq \mathbb{E}_g\left[\left|\int_{-k}^{k}\frac{1}{|\mathcal{M}_c[\varphi](t)|^2}Y_1^{c-1+2\pi it}(\overline{\mathcal{M}_c[g]}(t)-\overline{\mathcal{M}_c[g_o]}(t))\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right]\end{aligned}$$

With Lemma 1.2.5 (i) we obtain

$$\begin{aligned}&\mathbb{E}_g\left[\left|\int_{-k}^{k}\frac{1}{|\mathcal{M}_c[\varphi](t)|^2}Y_1^{c-1+2\pi it}(\overline{\mathcal{M}_c[g]}(t)-\overline{\mathcal{M}_c[g_o]}(t))\,\mathrm{w}^2(t)d\lambda(t)\right|^2\right]\\ &\leq \mathrm{C}_\varphi \mathrm{v}_{f|\varphi}\|(\mathcal{M}_c[g]-\mathcal{M}_c[g_o])\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^4)}.\end{aligned}$$

Exploiting $ab \leq a^2/4 + b^2$ for $a, b \in \mathbb{R}$, we get

$$\begin{aligned}
&\gamma^{-1/2}(\mathrm{C}_{\varphi}\mathrm{v}_{f|\varphi})^{1/2}\|(\mathcal{M}_c[g] - \mathcal{M}_c[g_o])\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^2(\mathrm{w}^4)} n^{-1/2} \\
&\leq \gamma^{-1/2}(\mathrm{C}_{\varphi}\mathrm{v}_{f|\varphi})^{1/2}\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^\infty(\mathrm{w})}\|(\mathcal{M}_c[g] - \mathcal{M}_c[g_o])\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|_{\mathbb{L}^2(\mathrm{w}^2)} n^{-1/2} \\
&\leq \frac{1}{4}\|(\mathcal{M}_c[g] - \mathcal{M}_c[g_o])\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^2(\mathrm{w}^2)} + \gamma^{-1}n^{-1}\mathrm{C}_{\varphi}\mathrm{v}_{f|\varphi}\|\mathbb{1}_{[-k,k]}/\mathcal{M}_c[\varphi]\|^2_{\mathbb{L}^\infty(\mathrm{w})} = \tau^{W_k}(\gamma)
\end{aligned}$$

and, consequently,

$$n\mathbb{E}^n_g[|W_k|^2] \leq n\gamma|\tau^{W_k}(\gamma)|^2.$$

Finally, by applying Markov's inequality follows immediately the result, that is,

$$\mathbb{P}^n_{\varphi,f}(W_k \geq \tau^{W_k}(\gamma)) \leq \frac{\mathbb{E}^n_g[|W_k|^2]}{|\tau^{W_k}(\gamma)|^2} \leq \gamma$$

which completes the proof. □

Perspectives

We give a short outlook on possible extensions for the questions discussed in this thesis, which might be addressed in future work.

Lower bounds In general, it is of interest to show minimax optimality, see Tsybakov (2008). Consequently, corresponding lower bounds need to be established for all upper bounds derived in this thesis. We expect the presented approaches to be minimax optimal, since they match standard rates for comparable settings. For example, the results of Chapter 2 correspond to upper bounds of direct density estimation with updated sample size, which are established to be minimax optimal, see for example Tsybakov (2008). For the multiplicative measurement error model considered in Chapters 3 and 4, rates of convergence for corresponding inference problems are proven to be minimax optimal, for example in the additive measurement error model, see Butucea (2004), or the circular convolution model, Schluttenhofer and Johannes (2020b). However, the derivation is left for future work.

Dependent data Throughout this thesis, we assumed observations of $(\boldsymbol{X}^{(j)})_{j\in[\![p]\!]}$ in Chapter 2 and observations Y in Chapters 3 and 4, to be independent, respectively. In some applications this might not be realistic. For both observation models considered, it is of interest to consider the case of dependence. However, for the techniques of this thesis the application of the convolution theorem is crucial. For this, independence is necessary. For the model presented in Chapter 2, an alternative might be to consider Copula-based convolution, as introduced in Cherubini et al. (2011) for two independent $\mathbb{R}$ valued random variables.

Perspectives for Chapter 2

We considered the estimation of the m-fold additive convolution of m-independent random vectors. We first give possible extensions for this setting.

Plug-in projection estimator In this scenario, consider a plug-in projection estimator analogously to the kernel plug-in estimator discussed in Section 2.1. For example, setting $d = 1$, given an orthonormal basis $(\varphi_\ell)_{\ell\in\mathbb{N}}$ of $\mathbb{R}$, consider for each component density f_j a projection estimator

$$\hat{f}_j := \sum_{\ell=0}^{m-1} \widehat{a}_\ell \varphi_\ell, \quad \widehat{a}_\ell := \frac{1}{n} \sum_{k\in[\![n]\!]} \varphi_\ell(X_k^{(1)}).$$

Then, propose the estimator consisting of the convolution of projection estimators for each component density, i.e.,

$$\widehat{g_{\boldsymbol{h}}^{(p)}}(\boldsymbol{x}) := \hat{f}_1 * \cdots * \hat{f}_p.$$

A problem with this approach is that the convolution of basis functions not necessarily results in basis functions again. For choices of basis other than the Hermite basis this might work. However, the proofs in this thesis heavily depend on properties of the Hermite basis, consequently, other techniques have to be applied.

Measurement errors In Chapter 2, we considered direct observations. In practice, observations are often not directly observed but with some measurement error. For example, one could assume to have access to independent samples of independent vectors, where each vector is deteriorated by some (known) additive measurement error. More precisely, assume in this scenario that we observe $p \in \mathbb{N}$ independent samples of independent vectors of dimension $d \in \mathbb{N}$ with possibly different sample sizes $n_j \in \mathbb{N}$ for all $j \in [\![p]\!]$, i.e.

$$\boldsymbol{Y}_k^{(j)} = \boldsymbol{X}_k^{(j)} + \boldsymbol{U}_k^{(j)}, \quad k \in [\![n_j]\!], j \in [\![p]\!],$$

where $\boldsymbol{X}_k^{(j)} \sim f_j$ and the known error $\boldsymbol{U}_k^{(j)} \sim \varphi_j$ are assumed to be independent. Analogously to Chapter 2, for estimating the additive convolution of the densities, i.e., $g^{(p)} := f_1 * \cdots * f_p$, consider the plug-in estimator using for each component density f_j an deconvolution approach, as in Meister (2009). That is, define

$$\widehat{g_{\boldsymbol{h}}^{(p)}}(\boldsymbol{x}) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle \boldsymbol{t}, \boldsymbol{x}\rangle} \mathcal{F}[L](\boldsymbol{ht}) \prod_{j \in [\![p]\!]} \widehat{\mathcal{F}[f_j]}(\boldsymbol{t}) d\lambda^d(\boldsymbol{t}).$$

with

$$\widehat{\mathcal{F}[f_j]}(\boldsymbol{t}) := \frac{1}{n_j} \sum_{k=1}^{n} \frac{e^{i\langle \boldsymbol{t}, \boldsymbol{Y}_k^{(j)}\rangle}}{\mathcal{F}[\varphi_j](\boldsymbol{t})}.$$

In practice the error distribution is unknown. Consequently, an extension to this case is of interest as well.

Estimating the multiplicative convolution of independent random variables Analogously to Chapter 2, a further question of interest is to consider for two positive real-valued independent random variables $X^{(1)} \sim f_1$ and $X^{(2)} \sim f_2$ the estimation of the density G of their multiplication, i.e., $X^{(1)} \cdot X^{(2)} \sim G$. In this case, G is given by the multiplicative convolution of the densities, that is, $x \in \mathbb{R}_+$

$$G(x) := f_1 \circledast f_2(x) := \int_{\mathbb{R}_+} f_1(y) f_2(x/y) y^{-1} d\lambda(y).$$

For the estimation of the multiplicative convolution of the densities, it is possible to propose an analogous plug-in estimator based on correspondingly adapting the Kernel estimator, see Brenner Miguel (2023) p.32. More precisely, for $x \in \mathbb{R}_+$

$$\hat{G}_h(x) := \frac{1}{n^2} \sum_{k_1 \in [\![n]\!]} \sum_{k_2 \in [\![n]\!]} L_h(x/(X_{k_1}^{(1)} X_{k_2}^{(2)})).$$

The analysis of this estimator then relies on proving corresponding properties of the multiplicative convolution. For example, note that Lemma 2.1.2 also holds for the multiplicative convolution.

Perspectives for Chapters 3 and 4

In this thesis, we considered (weighted) quadratic functional estimation and goodness-of-fit testing for the multiplicative measurement error model. The following presented extension may be addressed in a future work.

Unknown error distribution Throughout this thesis, the error distribution was assumed to be known, which might not be the case in practice. As discussed by Brenner Miguel (2023), without additional knowledge we cannot distinguish between signal and error. In this work, two scenarios are discussed: a second sample approach and panel data.

In the second sample scenario, we assume to have access additionally to the sample $(Y_j)_{j\in[\![n]\!]}$ to a second sample $(\tilde{U}_j)_{j\in[\![m]\!]}$ from the error variable U. Based on this sample, derive an estimator $\widetilde{\mathcal{M}}_c[\varphi]$ of the Mellin transform of the error density φ. Plugging this estimator into the quadratic functional estimator proposed in Equation (3.4) leads to estimator

$$\tilde{q}_k^2 := \frac{1}{n(n-1)} \sum_{\substack{j\neq l \\ j,l\in[\![n]\!]}} \int_{-k}^{k} \frac{Y_j^{c-1+2\pi it} Y_l^{c-1-2\pi it}}{|\widetilde{\mathcal{M}}_c[\varphi](t)|^2} \mathrm{w}^2(t) d\lambda(t).$$

Analogously, it is possible to adapt the test statistic $\hat{q}_k^2$ proposed in Equation (4.7). This approach has been already considered in Brenner Miguel et al. (2024) for density estimation in the multiplicative measurement error model.

Alternatively, Brenner Miguel (2023) proposes a panel data approach. Here, we assume to observe an i.i.d. sample $(\boldsymbol{Y}_j)_{j\in[\![n]\!]}$ of the random vector

$$\boldsymbol{Y} = (Y^{(1)}, Y^{(2)}, \dots, Y^{(m)}), \quad Y^{(j)} = X \cdot U^{(j)},$$

where $(U^{(j)})_{j\in[\![m]\!]}$ are i.i.d. In other words, we measure X m-times only differing in the measurement errors, leading to observations $Y^{(1)}, \dots, Y^{(m)}$. Inspired by Comte and Kappus (2015); Kappus and Mabon (2014) it is possible to deduce a representation of the Mellin transform of X in terms of the Mellin transform of Y which can be used to deduce an estimator of the quadratic functional.

Bibliography

Ahmad, I. and Fan, Y. (2001). Optimal bandwidths for kernel density estimators of functions of observations. *Statistics & Probability Letters*, 51(3):245–251.

Ahmad, I. A. and Mugdadi, A. R. (2003). Analysis of kernel density estimation of functions of random variables. *Journal of Nonparametric Statistics*, 15(4–5):579–605.

Ammous, S., Dedecker, J., and Duval, C. (2024). Adaptive directional estimator of the density in $\mathbb{R}^d$ for independent and mixing sequences. *Journal of Multivariate Analysis*, 203:105332.

Andersen, K. E. and Hansen, M. B. (2001). Multiplicative censoring: density estimation by a series expansion approach. *Journal of Statistical Planning and Inference*, 98(1):137–155.

Askey, R. and Wainger, S. (1965). Mean convergence of expansions in Laguerre and Hermite series. *American Journal of Mathematics*, 87(3):695–708.

Bandyopadhyay, S. (2012). A note on efficient density estimators of convolutions. *Journal of Statistical Planning and Inference*, 142(11):3056–3060.

Baraud, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606.

Baraud, Y., Huet, S., and Laurent, B. (2003). Adaptive tests of linear hypotheses by model selection. *The Annals of Statistics*, 31(1).

Barron, A., Birgé, L., and Massart, P. (1999). Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, 113(3):301–413.

Baudry, J.-P., Maugis, C., and Michel, B. (2012). Slope heuristics: Overview and implementation. *Statistics and Computing*, 22(2):455–470.

Belomestny, D., Comte, F., and Genon-Catalot, V. (2016). Nonparametric Laguerre estimation in the multiplicative censoring model. *Electronic Journal of Statistics*, 10(2):3114 – 3152.

Belomestny, D., Comte, F., and Genon-Catalot, V. (2019). Sobolev-Hermite versus Sobolev nonparametric density estimation on $\mathbb{R}$. *Annals of the Institute of Statistical Mathematics*, 71(1):29–62.

Belomestny, D. and Goldenshluger, A. (2020). Nonparametric density estimation from observations with multiplicative measurement errors. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 56(1):36 – 67.

Belomestny, D. and Goldenshluger, A. (2021). Density deconvolution under general assumptions on the distribution of measurement errors. *The Annals of Statistics*, 49(2):615 – 649.

Belomestny, D., Morozova, E., and Panov, V. (2024). Statistical inference for scale mixture models via mellin transform approach. *Statistics*, 58(1):209–229.

Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā, Series A*, 50(3):381–393.

Birgé, L. and Massart, P. (1995). Estimation of integral functionals of a density. *The Annals of Statistics*, 23(1):11 – 29.

Bissantz, N., Holzmann, H., and Proksch, K. (2014). Confidence regions for images observed under the Radon transform. *Journal of Multivariate Analysis*, 128:86–107.

Bongioanni, B. and Torrea, J. L. (2006). Sobolev spaces associated to the harmonic oscillator. *Indian Academy of Sciences. Proceedings. Mathematical Sciences*, 116(3):337–360.

Brenner Miguel, S. (2022). Anisotropic spectral cut-off estimation under multiplicative measurement errors. *Journal of Multivariate Analysis*, 190:104990.

Brenner Miguel, S. (2023). *The Mellin transform in nonparametric statistics*. Universitätsbibliothek Heidelberg, Heidelberg.

Brenner Miguel, S., Comte, F., and Johannes, J. (2021). Spectral cut-off regularisation for density estimation under multiplicative measurement errors. *Electronic Journal of Statistics*, 15(1):3551 – 3573.

Brenner Miguel, S., Comte, F., and Johannes, J. (2023). Linear functional estimation under multiplicative measurement error. *Bernoulli*, 29(3):2247 – 2271.

Brenner Miguel, S., Johannes, J., and Siebel, M. (2024). Multiplicative deconvolution under unknown error distribution. *Electronic Journal of Statistics*, 18(2):4795–4850.

Brenner Miguel, S. and Phandoidaen, N. (2022). Multiplicative deconvolution in survival analysis under dependency. *Statistics*, 56(2):297–328.

Brunel, E., Comte, F., and Genon-Catalot, V. (2016). Nonparametric density and survival function estimation in the multiplicative censoring model. *Test*, 25:570–590.

Butucea, C. (2001). Exact adaptive pointwise estimation on Sobolev classes of densities. *European Series in Applied and Industrial Mathematics (ESAIM): Probability and Statistics*, 5:1–31.

Butucea, C. (2004). Asymptotic normality of the integrated square error of a density estimator in the convolution model. *Sort: Statistics and Operations Research Transactions, Volume 28, Nº 1, 2004, pags. 9-26*, 28.

Butucea, C. (2007). Goodness-of-fit testing and quadratic functional estimation from indirect observations. *The Annals of Statistics*, 35(5):1907 – 1930.

Butucea, C. and Comte, F. (2009). Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli*, 15(1):69–98.

Butucea, C. and Meziani, K. (2011). Quadratic functional estimation in inverse problems. *Statistical Methodology*, 8(1):31–41. Advances in Data Mining and Statistical Learning.

Cavalier, L. (2000). Efficient estimation of a density in a problem of tomography. *The Annals of Statistics*, 28(2):630–647.

Cavalier, L. (2001). On the problem of local adaptive estimation in tomography. *Bernoulli*, 7(1):63–78.

Cherubini, U., Mulinacci, S., and Romagnoli, S. (2011). On the distribution of the (un)bounded sum of random variables. *Insurance: Mathematics and Economics*, 48(1):56–63.

Chesneau, C. (2011). On adaptive wavelet estimation of a quadratic functional from a deconvolution problem. *Annals of the Institute of Statistical Mathematics*, 63:405–429.

Chesneau, C., Comte, F., and Navarro, F. (2013). Fast nonparametric estimation for convolutions of densities. *Canadian Journal of Statistics*, 41(4):617–636.

Cohen, A., Davenport, M. A., and Leviatan, D. (2013). On the stability and accuracy of least squares approximations. *Foundations of Computational Mathematics*, 13(5):819–834.

Collier, O., Comminges, L., and Tsybakov, A. B. (2017). Minimax estimation of linear and quadratic functionals on sparsity classes. *The Annals of Statistics*, 45(3).

Comte, F. (2017). *Nonparametric Estimation.* Spartacus-Idh, Paris.

Comte, F. and Dion, C. (2016). Nonparametric estimation in a multiplicative censoring model with symmetric noise. *Journal of Nonparametric Statistics*, 28(4):768–801.

Comte, F., Duval, C., and Genon-Catalot, V. (2014). Nonparametric density estimation in compound Poisson processes using convolution power estimators. *Metrika*, 77(1):163–183.

Comte, F., Johannes, J., and Neubert, B. (2025). Quadratic functional estimation from observations with multiplicative measurement error. *Annals of the Institute of Statistical Mathematics.*

Comte, F. and Kappus, J. (2015). Density deconvolution from repeated measurements without symmetry assumption on the errors. *Journal of Multivariate Analysis*, 140:31–46.

Comte, F. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Annales de l'IHP Probabilités et statistiques*, 49:569–609.

Comte, F. and Lacour, C. (2023). Non compact estimation of the conditional density from direct or noisy data. *Annales de l'Institut Henri Poincaré. Probabilités et Statistiques*, 59(3):1463–1507.

Comte, F. and Marie, N. (2020). Bandwidth selection for the Wolverton–Wagner estimator. *Journal of Statistical Planning and Inference*, 207:198–214.

Comte, F. and Neubert, B. (2025). Estimation for the convolution of several multivariate densities. preprint.

Devroye, L. and Györfi, L. (1985). *Nonparametric density estimation. The L_1 view*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Hoboken, NJ.

Donoho, D. L. and Nussbaum, M. (1990). Minimax quadratic estimation of a quadratic functional. *Journal of Complexity*, 6(3):290–323.

Duval, C. (2013). Density estimation for compound Poisson processes from discrete data. *Stochastic Processes and their Applications*, 123(11):3963–3986.

Efromovich, S. (1997). Density estimation for the case of supersmooth measurement error. *Journal of the American Statistical Association*, 92(438):526–535.

Efromovich, S. (1999). *Nonparametric curve estimation. Methods, theory, and applications*. Springer Series in Statistics. New York, NY: Springer.

Engl, H. W., Hanke, M., and Neubauer, A. (1996). *Regularization of inverse problems*. Kluwer Academic Publishers, Dobrecht / Boston / London.

Ermakov, M. S. (1990). Minimax detection of a signal in a Gaussian white noise. *Theory of Probability and its Applications*, 35(4):667–679.

Ermakov, M. S. (1994). Minimax nonparametric testing of hypotheses on the distribution density. *Teoriya Veroyatnosteĭ i eë Primeneniya*, 39(3):488–512.

Frees, E. W. (1994). Estimating densities of functions of observations. *Journal of the American Statistical Association*, 89(426):517–525.

Fromont, M. and Laurent, B. (2006). Adaptive goodness-of-fit tests in a density model. *The Annals of Statistics*, 34(2).

Giné, E. and Mason, D. M. (2007a). Laws of the iterated logarithm for the local U-statistic process. *Journal of Theoretical Probability*, 20(3):457–485.

Giné, E. and Mason, D. M. (2007b). On local U-statistic processes and the estimation of densities of functions of several sample variables. *The Annals of Statistics*, 35(3).

Giné, E. and Nickl, R. (2016). *Mathematical foundations of infinite-dimensional statistical models*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, New York.

Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39(3):1608–1632.

Guo, H., Wang, J., and Tian, X. (2018). The mean consistency of wavelet estimators for convolutions of the density functions. *Journal of Computational and Applied Mathematics*, 343:1–11.

Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin*, pages 49–52.

Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*, volume 129 of *Lecture Notes in Statistics*. Berlin: Springer.

Hu, C., Pozdnyakov, V., and Yan, J. (2019). Density and distribution evaluation for convolution of independent gamma variables. *Computational Statistics*, 35(1):327–342.

Ingster, Y. and Suslina, I. A. (2012). *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169. Springer Science & Business Media, New York ; Berlin ; Heidelberg.

Ingster, Y. I. (1993a). Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Mathematical Methods of Statistics*, 2(2):85–114.

Ingster, Y. I. (1993b). Asymptotically minimax hypothesis testing for nonparametric alternatives. II. *Mathematical Methods of Statistics*, 2(3):171–189.

Ingster, Y. I. (1993c). Asymptotically minimax hypothesis testing for nonparametric alternatives. III. *Mathematical Methods of Statistics*, 2(4):249–268.

Kappus, J. and Mabon, G. (2014). Adaptive density estimation in deconvolution problems with unknown error distribution. *Electronic Journal of Statistics*, 8(2):2879–2904.

Korostelev, A. P. and Tsybakov, A. B. (1993). *Minimax theory of image reconstruction*, volume 82 of *Lecture Notes in Statistics*. New York, NY: Springer.

Kroll, M. (2019). Rate optimal estimation of quadratic functionals in inverse problems with partially unknown operator and application to testing problems. *ESAIM: Probability and Statistics*, 23:524–551.

Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *Comptes rendus de l'Académie des sciences. Série I, Mathématique*, 342(11):877–882.

Lacour, C., Massart, P., and Rivoirard, V. (2017). Estimator selection: a new method with applications to kernel density estimation. *Sankhyā, Series A*, 79:298–335.

Laurent, B. (1996). Efficient estimation of integral functionals of a density. *The Annals of Statistics*, 24(2):659–681.

Laurent, B. (2005). Adaptive estimation of a quadratic functional of a density by model selection. *ESAIM: Probability and Statistics*, 9:1–18.

Laurent, B., Loubes, J.-M., and Marteau, C. (2012). Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electronic Journal of Statistics*, 6(none).

Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302 – 1338.

Lee, A. J. (1990). *U-Statistics: Theory and Practice*. CRC Press, New York.

Lehmann, E. L. and Casella, G. (1998). *Theory of point estimation*. Springer Texts in Statistics. New York, NY: Springer, 2nd edition.

Lehmann, E. L. and Romano, J. P. (2005). *Testing statistical hypotheses.* Springer Texts Stat. New York, NY: Springer, 3rd edition.

Lepski, O. (1990). On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability and its Applications*, 35(3):454–466.

Lepski, O. (1991). Asymptotically minimax adaptive estimation. I: Upper bounds. Optimally adaptive estimates. *Theory of Probability and its Applications*, 36(4):682–697.

Lepski, O. (1992a). Asymptotically minimax adaptive estimation. II: Schemes without optimal adaption: Adaptive estimators. *Theory of Probability and its Applications*, 37(3):433–448.

Lepski, O. (1992b). On problems of adaptive estimation in white Gaussian noise. In *Topics in nonparametric estimation*, pages 87–106, Moscow/USSR. Advances in Soviet Mathematics.

Lepski, O. V. and Levit, B. Y. (1998). Adaptive minimax estimation of infinitely differentiable functions. *Mathematical Methods of Statistics*, 7(2):123–156.

Mabon, G. (2016). Adaptive deconvolution of linear functionals on the nonnegative real line. *Journal of Statistical Planning and Inference*, 178:1–23.

Marteau, C. and Sapatinas, T. (2017). Minimax goodness-of-fit testing in ill-posed inverse problems with partially unknown operators. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 53(4).

Massart, P. (2007). *Concentration inequalities and model selection. Ecole d'Eté de Probabilités de Saint-Flour XXXIII – 2003.* Lecture Notes in Mathematics (Volume 1896). Springer, Berlin.

Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics.* Springer, Berlin.

Midhu, N. N., Dewan, I., Sudheesh, K. K., and Sreedevi, E. P. (2023). On approximation and estimation of distribution function of sum of independent random variables. *Statistical Papers*, 65(3):1437–1467.

Nickl, R. (2007). Donsker-type theorems for nonparametric maximum likelihood estimators. *Probability Theory and Related Fields*, 138(3–4).

Nickl, R. (2009). On convergence and convolutions of random signed measures. *Journal of Theoretical Probability*, 22(1):38–56.

Panjer, H. H. and Willmot, G. E. (1992). *Insurance risk models.* Society of Acturaries, Schaumburg, IL.

Pensky, M. (2017). Minimax theory of estimation of linear functionals of the deconvolution density with or without sparsity. *The Annals of Statistics*, 45(4):1516–1541.

Prakasa Rao, B. L. S. (2004). Moment inequalities for supremum of empirical processes of U-statistic structure and application to density estimation. *Journal of the Iranian Statistical Society JIRSS*, 3(1):59–68.

Saavedra, A. and Cao, R. (2000). On the estimation of the marginal density of a moving average process. *Canadian Journal of Statistics*, 28(4):799–815.

Sacko, O. B. (2020). Hermite density deconvolution. *ALEA : Latin American Journal of Probability and Mathematical Statistics*, 17:419–443.

Schick, A. and Wefelmeyer, W. (2004a). Root n consistent and optimal density estimators for moving average processes. *Scandinavian Journal of Statistics*, 31(1):63–78.

Schick, A. and Wefelmeyer, W. (2004b). Root n consistent density estimators for sums of independent random variables. *Journal of Nonparametric Statistics*, 16(6):925–935.

Schick, A. and Wefelmeyer, W. (2007). Root-n consistent density estimators of convolutions in weighted L_1-norms. *Journal of Statistical Planning and Inference*, 137(6):1765–1774.

Schick, A. and Wefelmeyer, W. (2008). Root-n consistency in weighted L_1-spaces for density estimators of invertible linear processes. *Statistical Inference for Stochastic Processes*, 11(3):281–310.

Schick, A. and Wefelmeyer, W. (2009a). Convergence rates of density estimators for sums of powers of observations. *Metrika*, 69(2–3):249–264.

Schick, A. and Wefelmeyer, W. (2009b). Non-standard behavior of density estimators for sums of squared observations. *Statistics & Decisions*, 27(1):55–73.

Schick, A. and Wefelmeyer, W. (2012). On efficient estimation of densities for sums of squared observations. *Statistics & Probability Letters*, 82(9):1637–1640.

Schluttenhofer, S. (2020). *Adaptive minimax testing for inverse problems.* Universitätsbibliothek Heidelberg, Heidelberg.

Schluttenhofer, S. and Johannes, J. (2020a). Adaptive minimax testing for circular convolution. *Mathematical Methods of Statistics*, 29(2):106–133.

Schluttenhofer, S. and Johannes, J. (2020b). Minimax testing and quadratic functional estimation for circular convolution. Discussion paper.

Taupin, M.-L. (2001). Semi-parametric estimation in the nonlinear structural errors-in-variables model. *The Annals of Statistics*, 29(1):66–93.

Tsybakov, A. (2008). *Introduction to Nonparametric Estimation.* Springer Series in Statistics. New York, NY: Springer.

van der Vaart, A. and Wellner, J. A. (1996). *Weak convergence and empirical processes. With applications to statistics.* Springer Series in Statistics. New York, NY: Springer.

van derVaart, A. W. (2000). *Asymptotic statistics.* Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge, 1. paperback edition.

van Es, B., Klaassen, C. A., and Oudshoorn, K. (2000). Survival analysis under cross-sectional sampling: Length bias and multiplicative censoring. *Journal of Statistical Planning and Inference*, 91(2):295–312. Perspectives in modern statistical inference: parametric semiparametrics, non-parametrics.

Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika*, 76(4):751–761.

Vardi, Y. and Zhang, C.-H. (1992). Large sample study of empirical distributions in a random-multiplicative censoring model. *The Annals of Statistics*, 20(2):1022 – 1039.

Wasserman, L. (2006). *All of nonparametric statistics.* Springer Texts Stat. New York, NY: Springer.

Werner, D. (2018). *Funktionalanalysis.* Springer-Lehrbuch. Springer Spektrum, Berlin ; Heidelberg, 8. edition.

Willmot, G. E. and Woo, J.-K. (2007). On the class of erlang mixtures with risk theoretic applications. *North American Actuarial Journal*, 11(2):99–115.