INAUGURAL-DISSERTATION

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

vorgelegt von

Sergio Filipe Brenner Miguel
M.Sc. Mathematik

aus Heidelberg, Deutschland

Tag der mündlichen Prüfung:

The Mellin transform in Nonparametric Statistics

Betreuer: Prof. Dr. Jan Johannes

Acknowledgments

First of all, I am grateful to my supervisor Jan Johannes for his support, his time and guidance throughout the challenging years of my PhD. He always had time for a cup of coffee and a cigarette in his open office where we had the most interesting discussions.
I would also like to thank all of my collegues from Heidelberg and Mannheim who, especially during the Covid-19 pandemic, were a constant safe heaven from my worries and stress.

Secondly, I wish to express my thank Fabienne Comte for her time and advice during my research stay in Paris and outside of it. She, the professors and the PhD students from MAP5 made my Paris stay unforgettable. Merci pour votre patience avec mon français et soyez surs que vous restez un part très important dans mon histoire.

In particular, I thank Nathawut who was always there for me and supported me in and outside of work and Mariem who became in just a few months an important person in my life.

I am also deeply grateful to my family, Marion, Herminio, Andreas, Sara, Ana and Nuno who constantly reminded myself that there exist a life outside of work and my friends and loved ones Marcel, Lucia, Joos, Katharina, Stephanie, Maren, Can, Ido, Louis, Richard, Joachim and many more who were always there when I needed them the most. And thanks to Pedro who celebrated my frist publication with me with a bottle of sparkling wine.

Further, I would like to thank my lectors Nathawut, Bianca, Maxi and Richard for proofreading and helping to reduce the number of typos.

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 programme "Statistical Modeling of Complex Systems and Processes" grant RTG 1953, is also gratefully acknowledged.

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit der nichtparametrischen Schätzung für eine spezielle Klasse von schlecht-gestellten inversen Problemen, den sogenannten multiplikativen Messfehler-Modellen. In diesen Modellen sind die Beobachtungen der unbekannten, zu schätzenden Größe lediglich mit einem multiplikativen Messfehler zugänglich. Eine Konsequenz ist die Abhängigkeit der Stabilität der Rekonstruktion von der Verteilung der Fehlergröße, wobei diese maßgeblich die Schlechtgestelltheit des zugrundeliegenden inversen Problems beeinflusst. Die Theorie der Mellin-Transformierten ermöglicht zum Einen den Einfluss der Verteilung der Fehlergröße auf die Stabilität der Rekonstruktion zu beschreiben und zum Anderen die Schätzung der unbekannten Größen auf eine regularisierte Schätzung der unbekannten Mellin-Transformierten zurückzuführen. Die vorgestellten Schätzmethoden werden mittels eines zu erwartenden gewichtet(-integrierten) quadratischen Risikos evaluiert.
Neben einer Einführung in die Theorie der Mellin-Transfomierten und multiplikativen Faltungen ist die Arbeit in drei Themengebiete unterteilt.
Im ersten Teil betrachten wir die globale Dichteschätzung bei vorliegendem multiplikativen Messfehler. Nach einem Vergleich zwischen direkten und verrauschten Beobachtungen behandeln wir verschiedene Fehlerdichtenfamilien, den multivariaten Fall und den Einfluss von Abhängigskeitsstrukturen in den Daten. Hierbei wird jeweils eine Schätzmethode präsentiert, ihre Minimax-Optimalität diskutiert und daten-getriebene Wahlen des Glättungsparameters betrachtet. Mithilfe einer Monte-Carlo Simulation wird das theoretisch zu erwartende Verhalten des Schätzer illustriert.
Der zweite Teil der Arbeit beschäftigt sich mit der globalen Schätzung der Überlebensfunktion, die neben der Dichte einer Verteilung, eine gängige Charakterisierung von Verteilungen darstellt. Auch hier wird eine Schätzmethode vorgestellt, ihre Minimax-Optimalität bewiesen und daten-getriebene Wahlen von Glättungsparameter diskutiert. Des Weiteren wird die Stabilität des Schätzers für Bernoulli-Shift Prozesse analysiert und mittels einer Monte-Carlo Simulation visualisiert.
Der dritte Teil betrachtet die Schätzung der Evaluation eines linearen Funktionals im multiplikativen Messfehler-Model. Als Evaluation eines linearen Funktional können, unter anderem, die Punktevaluationen der Dichte, der Überlebungsfunktion und der Verteilungsfunktion interpretiert werden. Dies erlaubt es, diese unterschiedlichen Schätzprobleme simultan zu analysieren und Vergleiche bezüglich der Schlechtgestelltheit der inversen Probleme zu betrachten. Eine minimax-optimale Schätzmethode als auch eine daten-getriebene Wahl des Glättungsparameters werden präsentiert und analysiert.

Abstract

This thesis deals with the nonparametric estimation for a special class of ill-posed inverse problems, the so-called multiplicative measurement error models. In these models, the observations of the unknown, to be estimated quantity is only accessible with a multiplicative measurement error. As a consequence, the instability of the reconstruction depends on the distribution of the error by effecting the ill-posedness of the underlying inverse problem. The theory of Mellin transform allows to express the influence of the error distribution on the instability of the reconstruction and to reduce the estimation of the unknown quantity to a regularized estimation of its unknown Mellin transform. The proposed estimation strategies will be evaluated in terms of a mean weighted(-integrated) squared risk.

Aside from being an introduction to the theory of Mellin transforms and multiplicative convolutions, this thesis is structuered in three topics.

In the first part, we consider global density estimation under multiplicative measurement error. After a comparison between direct and noisy observations, we study several families of error distributions, the multivariate case and the influence of dependence structures in the data. Here in each case we will propose an estimation strategy, discuss its minimax-optimality and consider data-driven choices of smoothing parameters. The theoritcal expected behavior of the estimators are illustrated through Monte-Carlo simulations.

In the second part, we study global survival function estimation, which is, alongside the density of a distribution, a frequently considered characterization of a distribution. We once again propose an estimation method, prove its minimax-optimality and discuss data-driven choices of smoothing parameters. Furthermore, we analyse the stability of the estimator for Bernoulli-shift processes and visualize it using a Monte-Carlo simulation.

The third part considers the estimation of the evaluation of an linear functional under multiplicative measurement errors. The point evaluation of the density, the survival function and the cumulative distribution function, to mention only a few, can be intrepreted as an evaluation of a linear functional. This allows the simultaneous analysis of these different estimation problems and the comparison of the ill-posedness of the underlying inverse problems. A minimax-optimal estimation strategy as well as a data-driven choice of the smoothing parameters are presented and analyzed.

Contents

CHAPTER 1

Introduction

In this thesis, we are in general interested in the nonparametric estimation in a multiplicative measurement errors model. Starting with an motivation for the multiplicative measurement errors models, we will continue with an introduction to the nonparametric estimation framework and include the multiplicative measurement errors in the theory of inverse problems. We will then procede to a historical overview of the Mellin transform followed by a list of the contribution of this thesis to the existing nonparametric statistics literature.

Motivation

In the statistical inference, we are in general interested in using empirical data, for instance a sample $X_1, \dots, X_n$, to gain information about the underlying distribution $\mathbb{P}^X$. For example we are interested in the expected numbers of a cells on a 3D printed grid. This elementary problem can directly be solved by using the empirical mean of the sample $X_1, \dots, X_n$.
In practice however, the sample $X_1, \dots, X_n$ is often corrupted in the sense, that we might not be able to observe the number of cells explicitly. For example, cell counting programs sometimes count two cells which are too close to each other as one cell. This measurement error would always leads to a smaller number of cells per experiment. It is clear to see, that an additive model does not fit the described measurement error as it is of a ratio nature. Moreover, it is natural to model the measurement error as a random variable as it possibly depends on different latent effects.
In other words, instead of a sample $X_1, \dots, X_n$, we do only observe a sample $Y_1, \dots, Y_n$ where

$$Y_i = X_i \cdot U_i, \quad i = 1, \dots, n,$$

where the error term U_i is a random variable with values in $(0, 1)$. In this case, if the numbers of cells X_i and the measurement error of U_i are independent, the estimation of the mean is rather straight forward. Another question, for example the cestimation of the probability that the numbers of cells is bigger than 100, i.e. $\mathbb{P}(X > 100)$, leads to a very interesting but complicated statistical inverse problems which will be discussed throughout this thesis.
Before we formally introduce the multiplicative measurement errors model, we shall briefly recapitulate some elementary concepts of nonparametric estimation.

Nonparametric estimation

Let us consider a statistical experiment $(\Omega, \mathcal{A}, \mathcal{P}_\Theta)$ where $(\Omega, \mathcal{A})$ is a measurable space and $\mathcal{P}_\Theta := (\mathbb{P}_\theta)_{\theta\in\Theta}$ is a family of probability measures on $(\Omega, \mathcal{A})$. If the parameter space Θ can be identified with a subset of $\mathbb{R}^k$, $k \in \mathbb{N}$, one usually speaks of a PARAMETRIC FRAMEWORK. In contrary, if such an identification does not exist, the framework is said to be NONPARAMETRIC, compare Tsybakov (2009).

In other words, in nonparametric statistics, we do not assume that the parameter space Θ can be expressed as a subset of a finite-dimensional space. For instance, a frequently considered example of Θ would be the set of all square-integrable Lebesgue densities on $\mathbb{R}$.

In nonparametric estimation, we are interested in estimating $\theta_o \in \Theta$, respectively $\gamma(\theta_o)$ for $\gamma : \Theta \to \Gamma$ measurable, based on an identically distributed sample $X_1, \dots, X_n \sim \mathbb{P}_{\theta_o}$. Given a measurable loss function $d : \Gamma\times\Gamma \to [0, \infty)$, we measure the accuracy of an estimator $\widehat{\gamma} := \widehat{\gamma}(X_1, \dots, X_n) : \Omega \to \Gamma$ through the expected risk

$$\mathcal{R}_n(\widehat{\gamma}, \gamma(\theta_o)) := \mathbb{E}^n_{\theta_0}(d(\widehat{\gamma}, \gamma(\theta_o)))$$

of the estimator $\widehat{\gamma}$ given $\theta_o \in \Theta$. Here, $\mathbb{E}^n_{\theta_0}$ denotes the expectation with respect to the joint distribution of $X_1, \dots, X_n$. An example for such a risk is given by the mean integrated squared error (MISE) where Γ is the space of all square-integrable functions on $\mathbb{R}$ and $d(f, g) = \int_{\mathbb{R}} |f(x) - g(x)|^2 d\lambda(x)$ for $f, g \in \Gamma$. In this context, λ denotes the Lebesgue measure on $\mathbb{R}$.

A fundamental concept of nonparametric estimation consists of the minimax theory for estimation problems. We will now briefly summarize the main concepts of this theory while a more precise study in each estimation scenario can be found throughout this thesis.

For a class $\mathcal{F} \subseteq \Theta$, we define the MAXIMAL RISK over the class $\mathcal{F}$ of the estimator $\widehat{\gamma}$ through

$$\mathcal{R}_n(\widehat{\gamma}, \mathcal{F}) := \sup_{\theta\in\mathcal{F}} \mathbb{E}^n_\theta(d(\widehat{\gamma}, \gamma(\theta))).$$

Usually, the class $\mathcal{F}$ consists of an ellipsoid and/or express a regularity condition on the parameter $\theta \in \Theta$. As typical for nonparametric estimation problems, the maximal risk $\mathcal{R}_n(\widehat{\gamma}, \mathcal{F})$ might not decay with a so-called parametric rate, that is $\sup_{n\in\mathbb{N}} n \cdot \mathcal{R}_n(\widehat{\gamma}, \mathcal{F}) = \infty$. To show that the rate of convergence is an intrinsic feature of the statistical experiment itself rather than being a consequence of the choice of the estimator, we study the MINIMAX RISK over the class $\mathcal{F}$ defined by

$$\mathcal{R}_n(\mathcal{F}) := \inf_{\widehat{\gamma}} \mathcal{R}_n(\widehat{\gamma}, \mathcal{F}) = \inf_{\widehat{\gamma}} \sup_{\theta\in\mathcal{F}} \mathbb{E}_\theta(d(\widehat{\gamma}, \gamma(\theta))).$$

where the infimum is taken over all estimators based on the sample $X_1, \dots, X_n$. If the maximal risk of an estimator $\widehat{\gamma}$ behaves similar as the minimax risk, we would call it MINIMAX OPTIMAL. More precisely, an estimator $\widehat{\gamma}$ is minimax-optimal if there exists a constant $C > 0$ such that for all $n \in \mathbb{N}$ holds

$$\mathcal{R}_n(\mathcal{F}) \leq \mathcal{R}_n(\widehat{\gamma}, \mathcal{F}) \leq C\mathcal{R}_n(\mathcal{F}).$$

The first inequality is a direct consequence of the definition of $\mathcal{R}_n(\mathcal{F})$ and $\mathcal{R}_n(\widehat{\gamma}, \mathcal{F})$. The study of the minimax risk and the minimax optimality of the proposed estimation strategy will be a reoccurring topic throughout this thesis. For classical results in context of density and regression function estimation, we refer to Tsybakov (2009) and Comte (2017).

For the sake of simplicity, let us restrict ourselves to the case of $(\Omega, \mathcal{A}) = (\mathbb{R}_+, \mathcal{B}_+)$ where $\mathbb{R}_+ := (0, \infty)$ and $\mathcal{B}_+$ denotes the Borel σ-field on $\mathbb{R}_+$. In practice, the observations $X_1, \dots, X_n$ are often only accessible in presence of measurement errors or censoring. Here, the additive noise model is a frequently considered noise model, see Meister (2009) for an overview of nonparametric estimation in additive noise models. In these model, we assume to observe identically distributed copies of Y with

$$Y = X + \varepsilon,$$

where ε is a second random variable, stochastically independent of X, the so-called error term. In this thesis, we will focus on multiplicative measurement errors, that is

$$Y = X \cdot U,$$

where U is second random variable, stochastically independent of X. A brief introduction and how these two models are connected will be discussed in the following.

Multiplicative measurement errors model

Formally, multiplicative measurement error models can be characterised by a statistical experiment $(\mathbb{R}_+, \mathcal{B}_+, \mathcal{P}_\Theta)$ where we only have access to identically distributed copies of

$$Y = X \cdot U, \quad X \sim \mathbb{P}_{\theta_o}, U \sim \mathbb{P}_U, \tag{1.1}$$

where X and U are stochastically independent. Our goal is to estimate θ_o, respectively $\gamma(\theta_o)$, given the observations $Y_1, \dots, Y_n$.
This general model was studied in Belomestny and Goldenshluger (2020) and includes a variety of frequently considered models as the multiplicative censoring model, see Vardi (1989), and the stochastic volatility model, see Genon-Catalot et al. (2003), Comte and Genon-Catalot (2006), to mention only a few.
Due to the independence of X and U the distribution $\mathbb{P}_Y$ of Y is given by the multiplicative convolution of the probability measures $\mathbb{P}_{\theta_o}$ and $\mathbb{P}_U$. Exploiting the Mellin transform, Belomestny and Goldenshluger (2020) show that the multiplicative convolution can be rewritten as the product of the Mellin transforms, that is $\mathcal{M}[\mathbb{P}_Y] = \mathcal{M}[\mathbb{P}_{\theta_o}] \cdot \mathcal{M}[\mathbb{P}_U]$. (The precise definition of the Mellin transform $\mathcal{M}[\mathbb{P}]$ of a probability measure on $(\mathbb{R}_+, \mathcal{B}_+)$ will be given in Chapter 2. For now, it suffices to read $\mathcal{M}[\mathbb{P}]$ as a bounded function $\mathcal{M}[\mathbb{P}] : \mathbb{R} \to \mathbb{C}$.)
In other words, the Mellin transform allows us to bring the convolution operator $\mathrm{T} : \mathcal{H} \to \mathcal{H}$, for a Hilbert space $\mathcal{H}$ to be specified later on, in a diagonal form, i.e.

$$\widetilde{\mathrm{T}} : \mathbb{L}^2(\mathbb{R}) \to \mathbb{L}^2(\mathbb{R}), \quad \widetilde{\mathrm{T}}h = h \cdot \mathcal{M}[\mathbb{P}_U] \tag{1.2}$$

where $\mathbb{L}^2(\mathbb{R})$ denotes the space of all square-integrable, complex-valued function on $\mathbb{R}$ endowed with the $\mathbb{L}^2$ inner product $\langle \cdot, \cdot \rangle_{\mathbb{L}^2(\mathbb{R})}$. Next, we clarify the terminology of inverse problems and then identify the estimation under multiplicative measurement errors as a statistical inverse problem.

Inverse problems

Given two separable Hilbert spaces $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ and $(\mathcal{G}, \langle \cdot, \cdot \rangle_{\mathcal{G}})$ and a linear operator

$$\mathrm{T} : \mathcal{H} \to \mathcal{G}, \quad h \mapsto \mathrm{T}h$$

recovering the element $h_o \in \mathcal{H}$ given a noisy observation of $g_o = \mathrm{T}h_o$ is called a STATISTICAL INVERSE PROBLEM. Naturally, the inverse of T, if it exists, characterizes the stability of the recovering of $h_o \in \mathcal{H}$. A statistical inverse problem is called WELL-POSED, compare Hadamard (1902), if the following three conditions hold

- EXISTENCE. There exists an element $h \in \mathcal{H}$ such that $\mathrm{T}h = g_o$, in other words $g_o \in \mathrm{T}\mathcal{H}$.
- UNIQUENESS. For all $h, h_o \in \mathcal{H}$ holds: $\mathrm{T}h = g = \mathrm{T}h_o$ implies $h = h_o$, in other words the operator T is injective.
- STABILITY. The inverse operator $\mathrm{T}^\dagger$ is continuous, that is for all $\varepsilon > 0$ there exists $\delta > 0$ such that $\|\mathrm{T}^\dagger g_1 - \mathrm{T}^\dagger g_2\|_{\mathcal{H}} \leq \varepsilon$ for all $g_1, g_2 \in \mathcal{G}$ with $\|g_1 - g_2\|_{\mathcal{G}} \leq \delta$.

In this thesis, the existence of a solution for the operator in (1.2) is equivalent with the fact that $g_o/\mathcal{M}[\mathbb{P}_U]\mathbb{1}_{\{\mathcal{M}[\mathbb{P}_U]\neq 0\}} \in \mathcal{H}$ which is guaranteed by the assumption that our data is generated by the equation (1.1). From (1.2) we see that the uniqueness of a solution is ensured as soon if the set of all zeros of the Mellin transform of $\mathbb{P}_U$ is a Lebesgue null-set which is a common assumption in the additive noise model, compare Meister (2009). Furthermore, we will discuss that in terms of stability, we see that the inverse operator is not continuous as soon as the Mellin transform $\mathcal{M}[\mathbb{P}_U](t)$ tends to zero for $|t|$ going to infinity. This is already the case if U possesses a Lebesgue density. It is therefore that the estimation under multiplicative measurement errors is in many cases an ILL-POSED statistical inverse problem.

In the case of ill-posed inverse problem, we cannot ensure that a good approximate of g_o would lead to a good approximate of h_o, or that even the noisy observation of g_o lies inside of the domain of $\mathrm{T}^\dagger$. To overcome this instability issue REGULARIZATION METHODS are needed. In this theses, we primary use the spectral cut-off regularization method, respectively a variation of it, which we present first in the setting of inverse problems and then in the particular case of multiplicative measurement error models.
Let $\mathrm{T} : \mathcal{H} \to \mathcal{H}$ be a normal operator. Then due to the spectral theorem of normal operators, compare Werner (2006), the normal operator is unitary equivalent to a multiplication operator. More precisely, there exists a measure space $(\Omega, \mathscr{A}, \mu)$, an unitary operator $\mathrm{U} : \mathcal{H} \to \mathbb{L}^2(\mu)$ and a bounded, complex-valued measurable function $\tau : \Omega \to \mathbb{C}$ such that

$$(\mathrm{UTU}^\dagger)\varphi = \tau \cdot \varphi =: \mathrm{M}_\tau \varphi, \quad \mu - \text{ almost everywhere}, \varphi \in \mathbb{L}^2(\mu). \tag{1.3}$$

In view of the spectral theorem of normal operators, it is natural to focus on the case of multiplication operators as the one expressed in (1.2). Thus, we assume from now on that $T = \mathrm{M}_\tau$, compare (1.3).

- SPECTRAL CUT-OFF REGULARIZATION Let $\mathbb{L}^2(\mu)$ be the $\mathbb{L}^2$ space with respect to a given measure space $(\Omega, \mathscr{A}, \mu)$ and

  $$\mathrm{T} : \mathbb{L}^2(\mu) \to \mathbb{L}^2(\mu), \; \varphi \mapsto \mathrm{T}\varphi := \tau \cdot \varphi,$$

  for a bounded, measurable function $\tau : \Omega \to \mathbb{C}$. The spectral cut-off regularized inverse of $\mathrm{T}^\dagger$ is then given by the family $(\mathrm{T}_k^\dagger)_{k \in \mathbb{R}_+}$. where

  $$\mathrm{T}_k^\dagger : \mathbb{L}^2(\mu) \to \mathbb{L}^2(\mu), \; \varphi \mapsto \mathrm{T}_k^\dagger \varphi := \frac{\varphi}{\tau} \mathbb{1}_{B_k}$$

where $(B_k)_{k\in\mathbb{R}_+}$ is a nested family of subsets of Ω such that $\bigcup_{k\in\mathbb{R}_+} B_k = \Omega$ and τ is bounded from below on each $B_k, k \in \mathbb{R}_+$.

In the case of the multiplicative convolution, see (1.2), the function τ is given through $\mathcal{M}[\mathbb{P}_U]$. Under some regularity assumption on $\mathcal{M}[\mathbb{P}_U]$, we then choose $B_k := [-k, k]$ for $k \in \mathbb{R}_+$. Given an estimator $\widehat{\mathcal{M}}[\mathbb{P}_Y]$ of $\mathcal{M}[\mathbb{P}_Y]$, the spectral cut-off regularization motivates a family of estimator $(\widehat{\mathcal{M}}_k[\mathbb{P}_{\theta_0}])_{k\in\mathbb{R}_+}$ of $\mathcal{M}[\mathbb{P}_{\theta_o}]$ given by

$$\widehat{\mathcal{M}}_k[\mathbb{P}_{\theta_0}] := \mathrm{T}_k^\dagger \widehat{\mathcal{M}}[\mathbb{P}_Y] = \frac{\widehat{\mathcal{M}}[\mathbb{P}_Y]}{\mathcal{M}[\mathbb{P}_U]} \mathbb{1}_{[-k,k]}$$

from which we will deduce a family of estimator $(\widehat{\gamma}_k)_{k\in\mathbb{R}_+}$, indexed by the cut-off parameter $k \in \mathbb{R}_+$, of the quantity of interest $\gamma(\theta_o)$. A choice of the cut-off parameter $k \in \mathbb{R}_+$ which minimizes the risk of the estimation of $\gamma(\theta_o)$ is in general non-trivial, as the dependency of the risk and the cut-off parameter is neither monotone nor known. Strategies to choose the parameter $k \in \mathbb{R}_+$ only depending on the observations are called data-driven method and represent next to the minimax theory an important field inside the nonparametric estimation literature.

Data-driven choices of smoothing parameters

In nonparametric estimation there exists many heuristics to construct estimators. Some examples are orthonormal series estimators, kernel estimators or estimators based on a regularization strategy of an inverse problem, compare Comte (2017) and Tsybakov (2009) for an overview. In general, these heuristics lead not to one single estimator but a family $(\widehat{\gamma}_k)_{k\in\mathcal{K}}$ of estimator of $\gamma(\theta_o)$ indexed by a smoothing parameter $k \in \mathcal{K}$ and a parameter space $\mathcal{K}$. For kernel estimators the smoothing parameter $k \in \mathcal{K} = \mathbb{R}_+$ is called bandwidth while for orthonormal series estimators $k \in \mathcal{K} = \mathbb{N}$ is referred to as dimension parameter. In most of these settings, the choice of the smoothing parameter is non-trivial. The existing literature addresses this issues by proposing and studying so-called DATA-DRIVEN OR ADAPTIVE methods. In the statistical literature, these two expressions are sometimes used synonymous.

To avoid confusion, we will only use the expression data-driven methods where we call any choice $\widehat{k} \in \mathcal{K}$ depending measurable only on the observations, a data-driven choice of the smoothing parameter. Some authors distinguish between PARTIALLY OR FULLY data-driven methods, if the choice is additionally dependent on some knowledge of the distribution of the observations or not. In this thesis, we only consider FULLY DATA-DRIVEN METHODS and call them shortly DATA-DRIVEN METHODS.

A desirable property of a data-driven method $\widehat{k}$ are risk bounds of the form

$$\mathcal{R}_n(\widehat{\gamma}_{\widehat{k}}, \gamma(\theta_o)) \leq C_1 \inf_{k\in\mathcal{K}} \mathcal{R}_n(\widehat{\gamma}_k, \gamma(\theta_o)) + C_2 r_n \tag{1.4}$$

where $C_1, C_2 \in \mathbb{R}_+$ are positive constants possibly depending on θ_o and r_n is a negligible rest term which vanishes if the sample size $n \in \mathbb{N}$ goes to infinity. In other words, a data-driven choice $\widehat{k}$ behaves, in terms of its risk, like the best choice of $k \in \mathcal{K}$ up to a multiplicative constant and a negligible additive rest term. In the literature, we often find bounds results of the form

$$\mathcal{R}_n(\widehat{\gamma}_{\widehat{k}}, \gamma(\theta_o)) \leq C_1 \inf_{k\in\mathcal{K}_n} \mathcal{R}_n(\widehat{\gamma}_k, \gamma(\theta_o)) + C_2 r_n \tag{1.5}$$

instead of (1.4) for a subset $\mathcal{K}_n \subseteq \mathcal{K}$. If $\inf_{k\in\mathcal{K}_n} \mathcal{R}_n(\widehat{\gamma}_k, \gamma(\theta_o)) \leq C_3 \inf_{k\in\mathcal{K}} \mathcal{R}_n(\widehat{\gamma}_k, \gamma(\theta_o))$ for a positive constant $C_3 \in \mathbb{R}_+$, then we say that $\mathcal{K}_n \subset \mathcal{K}$ is sufficiently large as it leads to no change of the rate. This may not always be the case.
Similar to the minimax theory, we will propose in each section a data-driven method for the choice of the upcoming smoothing parameters and prove upper boundaries of the form (1.5).
We present now a brief overview of the history of the Mellin transform in Mathematics and its role in Probability theory and Statistics.

History of the Mellin transform

The Mellin transform is an integral transform which is used in number theory, mathematical statistics and the theory of asymptotic expansions. It is named after the Finnish mathematician Hjalmar Mellin who was born in 1854 and died in 1933. During his studies Mellin was introduced to the field of complex analysis by his professor Gösta Mittag-Leffler. In his research he intensely studied the properties of the Γ-function and its relationship to linear differential and difference equations. One of his most famous results states that the conditions under which two functions F, Φ satisfy the relationship

$$\begin{cases} \Phi(x) = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} F(z) x^{-z} dz, \\ F(z) = \int_0^\infty \Phi(x) x^{z-1} dx \end{cases}$$

compare Lindelöf (1933).
In 1948, the mathematician Benjamin Epstein proposed in his paper, Epstein (1948), the usage of the Mellin transform in probability theory to determine the distribution of a product, respectively a quotient, of random variables. From there many authors have used the Mellin transform in their analysis of products of random variables, for instance Fox (1957), Wells et al. (1962), Springer and Thompson (1966), Lomnicki (1967), Subrahmaniam (1970) and Springer and Thompson (1970) to mention a few.
In the last years, the Mellin transform found its application in modern nonparametric statistics and was used by Belomestny and Schoenmakers (2016), Belomestny and Goldenshluger (2020) and Geenens (2021). Compared to the Fourier transform the Mellin transform has not been used frequently in nonparametric statistics although there exists several models involve the product of two independent random variables. To establish the usage of the Mellin transform for nonparametric statistics, we are in need of a theory of Mellin transforms similar to the functional analytical theory of Fourier transform, respectively the measure theoretical notion of characteristic functions which is one motivation for the underlying theses.

Contribution and structure of this thesis

This thesis is divided in five chapters. While the first two chapters present a general introduction in the underlying mathematical concepts and objects, chapter 3, 4 and 5 are dedicated to different quantities of interest which are estimated in a model with multiplicative measurement errors.
In general all chapters, respectively the sections of the chapter, begin with a short summary of the proposed results and related literature. Within the sections the main results and necessary definitions are displayed, discussed and illustrated by several examples. Proofs, which allows for a

deeper understanding of the topic, are presented within the sections while others are postponed to a dedicated proof subsection.
Let us summarize the main results of this thesis.

Chapter 2 - The Mellin transform and multiplicative convolution

In nonparametric Statistics, we often rely on functional analytical properties of the studied objects which are used for the construction of estimators. Although our focus is on the probabilistic respectively statistical aspects of the theory, a sufficiently large collection of properties of the considered, functional analytical objects is an essential part of the mathematical theory.
To our knowledge such a collection of properties of the Mellin transform does not exist as it has been rarely used in the context of nonparametric Statistics. It is therefore we collect and prove these properties in Section 2.
Our main contributions are the following

- We collect and prove properties of the multiplicative convolution from a functional analytical and probabilistic point of view.
- Further we define the Mellin transform in a functional analytical and probabilistic context and study their properties.
- We develop a canonical notion of multivariate Mellin transforms and their corresponding regularity classes.

Parts of this chapter has been published in Brenner Miguel et al. (2021), Brenner Miguel (2021), Brenner Miguel (2022a), Brenner Miguel (2022b), Brenner Miguel and Phandoidaen (2022) and Brenner Miguel et al. (2023).

Chapter 3 - Global density estimation under multiplicative measurement errors

In this chapter, we study the estimation of the density of a positive (multivariate) random variable with respect to a global risk. This chapter is divided into six sections each with varying assumptions on the error distribution, respectively the dependency structure of the observations.
In all sections, we contribute to the existing literature the following results

- We construct an estimator based on a regularization of the inverse Mellin transform.
- By proving upper and lower bounds of the risk, we establish the minimax optimality of the proposed estimator with respect to Mellin Sobolev spaces.
- We propose and study data-driven choices of the upcoming smoothing parameters.

Furthermore we briefly summarize the differences of the sections 3.1 to 3.6.

3.1 Direct observation: We study the case without multiplicative measurement error and introduce general strategies and concepts which are revisited in the following sections.

3.2 Multiplicative measurement errors: The multiplicative measurement error model is studied under the assumption of smooth error densities.

3.3 Super smooth error densities: The results proposed in 3.2 are shown under the assumption of super smooth error densities.

3.4 Oscillating densities: Another class of error densities, the so-called oscillating error densities are studied in this section.

3.5 Multivariate densities: While Section 3.1-3.4 consider univariate density estimation, the generalization of the estimation procedure to multivariate density estimation can be found in this section.

3.6 Stationary processes: Section 3.1-3.5 assume independency of the observations. Here, we only assume stationarity of the observation and study the resulting differences in the risk bounds of the proposed estimators. Furthermore, we address the volatility density estimation in a stochastic volatility model using the presented techniques of the chapter 3.

Section 3.1 and 3.2 has been published independently in Brenner Miguel et al. (2021), as well as Section 3.5 in Brenner Miguel (2021), Section 3.4 in Brenner Miguel (2022a) and Section 3.6 in Brenner Miguel (2022b).

Chapter 4 - Global survival function estimation under multiplicative measurement errors

With respect to a global risk, we study in this chapter the estimation of the survival function of a positive random variable under multiplicative measurement errors. Here, we restrict our studies to the case of smooth error densities and stationary processes. The chapter is separated in to sections considering the independent and the dependent case. In all sections, we contribute to the existing literature the following

- We construct an estimator based on the regularization of the inverse Mellin transform.
- By proving upper and lower bounds of the risk, we show the minimax optimality of the proposed estimator with respect to the Mellin Sobolev spaces
- We propose and study data-driven choices of the upcoming smoothing parameters.

The results of this section has been published independently in Brenner Miguel and Phandoidaen (2022).

Chapter 5 - Local estimation under multiplicative measurement errors

In this chapter, we consider the estimation of the evaluation of a linear functional with respect to the mean squared error. Since the point evaluation can be interpreted as an evaluation of a linear functional, this setting includes the estimation of several quantities of interests with respect to the point wise risk. For instance the point wise density and survival function estimation are included in this setting. We contribute to the existing literature the following

- We construct an estimator based on the regularization of the inverse Mellin transform.
- By proving upper and lower bounds of the risk, we show the minimax optimality of the proposed estimator with respect to the Mellin Sobolev spaces

- We propose and study data-driven choices of the upcoming smoothing parameters.

The results of this section has been published independently in Brenner Miguel et al. (2023).

List of publications

- Brenner Miguel, S. (2021). Multiplicative deconvolution based on a ridge approach, *arXiv preprint arXiv:2108.01523.*
- Brenner Miguel, S. (2022a). Anisotropic spectral cut-off estimation under multiplicative measurement errors *Journal of Multivariate Analysis,* 190:104990.
- Brenner Miguel, S. (2022b). Volatility density estimation by multiplicative deconvolution. *forthcoming in the: Journal of Nonparametric Statistics.*
- 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 errors, *forthcoming in: Bernoulli.*
- Brenner Miguel, S, Phandoidaen, N. (2022). Multiplicative deconvolution in survival analysis under dependency. *Statistics*, 56(2):297–328.

CHAPTER 2

The Mellin transform and multiplicative convolution

2.1 The Mellin transform for distributions

In this section, we start to define the multiplicative convolution and the Mellin transform for distributions similar to the definitions of additive convolution and characteristic functions. Inspired by the characteristic function chapter in Durrett (2019), we show corresponding results for the Mellin transform and the multiplicative convolution of distributions.

2.1.1 Multiplicative convolution

Let us consider the measurable space $(\mathbb{R}_+, \mathscr{B}_+)$ where we denote by $\mathbb{R}_+ := (0, \infty)$ the set of all strictly positive real numbers and $\mathscr{B}_+$ the Borel σ-algebra on $\mathbb{R}_+$.
Let X, Y be two independent $\mathbb{R}_+$ random variables with distribution $\mathbb{P}^X$ and $\mathbb{P}^Y$ on $(\mathbb{R}_+, \mathscr{B}_+)$. It is well-known that the random vector (X, Y) follows the product measure $\mathbb{P}^X \otimes \mathbb{P}^Y$ on $\mathbb{R}_+^2$. Therefore, the random variable XY follows the distribution $(\mathbb{P}^X \otimes \mathbb{P}^Y) \circ \odot^{-1}$, where $\odot : \mathbb{R}_+^2 \to \mathbb{R}_+, (x, y) \mapsto xy$ which leads to the following definition.

Definition 2.1.1 (Multiplicative Convolution)**:**
Let $\mathbb{P}, \mathbb{Q}$ be two probability measures on $(\mathbb{R}_+, \mathscr{B}_+)$. Then we call the probability measure

$$\mathbb{P} * \mathbb{Q} := (\mathbb{P} \otimes \mathbb{Q}) \circ \odot^{-1}$$

on $(\mathbb{R}_+, \mathscr{B}_+)$ the MULTIPLICATIVE CONVOLUTION of $\mathbb{P}$ and $\mathbb{Q}$.

By the definition of the multiplicative convolution its commutativity and associativity follows straightforward. If both measures are dominated by the counting measure, respectively the Lebesgue measure, we can express the multiplicative deconvolution in a more explicit form.

Proposition 2.1.2:
Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $(\mathbb{R}_+, \mathscr{B}_+)$.

(i) Let $A \subseteq \mathbb{R}_+$ be a discrete, or an at most countable, subset of $\mathbb{R}_+$ such that $A \odot A \subseteq A$. If the probability measures $\mathbb{P}$ and $\mathbb{Q}$ are dominated by the counting measure $\delta_{|A}$ on A with Radon-Nikodym derivatives $p := \frac{d\mathbb{P}}{d\delta_{|A}}$ and $q := \frac{d\mathbb{Q}}{d\delta_{|A}}$ then $\mathbb{P} * \mathbb{Q}$ is dominated by $\delta_{|A}$ and has for any $a \in A$ the Radon-Nikodym derivative

$$(p * q)(a) := \sum_{b \in A} p(a/b)q(b), \qquad \text{where } p(z) := 0 \text{ for } z \in \mathbb{R}_+ \setminus A.$$

(ii) If the probability measures $\mathbb{P}$ and $\mathbb{Q}$ are dominated by the Lebesgue measure λ with Radon-Nikodym derivatives $f = \frac{d\mathbb{P}}{d\lambda}$ and $g = \frac{d\mathbb{Q}}{d\lambda}$ then $\mathbb{P} * \mathbb{Q}$ is dominated by λ and has the Radon-Nikodym derivative

$$(f * g)(y) := \int_{\mathbb{R}_+} f(y/x)g(x)x^{-1}d\lambda(y), \qquad \text{for any } y \in \mathbb{R}_+.$$

Proof of Proposition 2.1.2. Let $A \subseteq \mathbb{R}_+$ with $A \odot A \subseteq A$ and A at most countable. For $a \in A$ we have $\odot^{-1}(a) = \bigcup_{b \in \mathbb{R}_+} \{(b, a/b)\}$. Since $\mathbb{P}$ is dominated by the measure $\delta_{|A}$ we get

$$(\mathbb{P} * \mathbb{Q})[\{a\}] = \sum_{b \in A} \mathbb{P} \otimes \mathbb{Q}[\{(b, a/b)\}] = \sum_{b \in A} p(b)q(a/b).$$

For the second part, let $y \in \mathbb{R}_+$. Then,

$$(\mathbb{P} * \mathbb{Q})[(-\infty, y)] = \int_{\mathbb{R}_+} \int_{(0,y/x)} f(x)g(z)d\lambda(z)d\lambda(x) = \int_{(0,y)} \int_{\mathbb{R}_+} f(x)g(z/x)x^{-1}d\lambda(x)d\lambda(z)$$

by application of the Fubini-Tonelli theorem and a change of variable. We deduce that $\mathbb{P}*\mathbb{Q}$ is dominated by the Lebesgue measure with Radon-Nikodym density $y \mapsto \int_{\mathbb{R}_+} f(x)g(y/x)x^{-1}d\lambda(x)$. □

Let us now illustrate the multiplicative convolution using the example of Log-Gamma and Log-Normal distributed random variables.

Example 2.1.3 (Log-Gamma Distribution):
The family of Log-Gamma densities $(f_{a,\sigma})_{(a,\sigma) \in \mathbb{R}_+^2}$ with respect to the Lebesgue measure is given by

$$f_{a,\sigma}(x) = \frac{\sigma^a}{\Gamma(a)} x^{-\sigma-1}(\log(x))^{a-1}\mathbb{1}_{(1,\infty)}(x) \text{ for } x \in \mathbb{R}_+.$$

Here Γ denotes the Gamma function $\Gamma : \mathbb{C} \setminus \mathbb{N}_0^- \to \mathbb{C}, s \mapsto \int_0^\infty x^{s-1} \exp(-x)dx$ with $\mathbb{N}_0^- := \{-n : n \in \mathbb{N}_0\}$. Then for arbitrary $\sigma, a_1, a_2 \in \mathbb{R}_+$,

$$f_{a_1,\sigma} * f_{a_2,\sigma} = f_{a_1+a_2,\sigma}.$$

This can be deduced from the formula in Proposition 2.1.2.

Example 2.1.4 (Log-Normal Distribution):
The family of Log-Normal densities $(f_{\mu,\sigma^2})_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_+}$ with respect to the Lebesgue measure is given by

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp(-(\log(x) - \mu)^2/2\sigma^2)\mathbb{1}_{\mathbb{R}_+}(x) \text{ for } x \in \mathbb{R}_+.$$

Then for any $\sigma_1^2, \sigma_2^2 \in \mathbb{R}_+$ and for any $\mu_1, \mu_2 \in \mathbb{R}$,

$$f_{\mu_1,\sigma_1^2} * f_{\mu_2,\sigma_2^2} = f_{\mu_1+\mu_2,\sigma_1^2+\sigma_2^2}.$$

This can be deduced from the formula in Proposition 2.1.2.

A more elegant way to show that the convolution properties in Example 2.1.3 and Example 2.1.4 is by using the theory of Mellin transforms which will be presented next.

2.1.2 The Mellin transform

Let us denote by $\mathcal{P}^+$ the set of all probability measures on $(\mathbb{R}_+, \mathscr{B}_+)$. For $c \in \mathbb{R}$ we set

$$\mathcal{P}_c^+ := \{\mathbb{P} \in \mathcal{P}^+ : \mathbb{E}_{\mathbb{P}}(X^{c-1}) := \int_{\mathbb{R}_+} x^{c-1} d\mathbb{P}(x) < \infty\},$$

as the set of all probability measures on $(\mathbb{R}_+, \mathscr{B}_+)$ with finite $c-1$ moment. Note that $\mathcal{P}^+ = \mathcal{P}_1^+$.

Definition 2.1.5 (Mellin transform for distributions):
Let $\mathbb{P} \in \mathcal{P}_c^+$ for a $c \in \mathbb{R}$. Then we define the Mellin transform $\mathcal{M}_c[\mathbb{P}]$ of the distribution $\mathbb{P}$ developed in the point c as the function $\mathcal{M}_c[\mathbb{P}] : \mathbb{R} \to \mathbb{C}$, by

$$\mathcal{M}_c[\mathbb{P}](t) := \mathbb{E}_{\mathbb{P}}(X^{c-1+it}) = \int_{\mathbb{R}_+} x^{c-1+it} d\mathbb{P}(x) \tag{2.1}$$

for any $t \in \mathbb{R}$. Here $i \in \mathbb{C}$ denotes the imaginary unit.

Given that $\mathbb{P}$ is dominated by the Lebesgue measure, respectively the counting measure on a discrete set $A \subseteq \mathbb{R}_+$, we can rewrite Equation (2.1) as

$$\mathcal{M}_c[\mathbb{P}](t) = \int_{\mathbb{R}_+} x^{c-1+it} f(x) d\lambda(x) \quad \text{resp.} \quad \mathcal{M}_c[\mathbb{P}](t) = \sum_{b\in A} b^{c-1+it} p(b)$$

for any $t \in \mathbb{R}$ where f and p are the Radon-Nikodym derivatives with respect to the Lebesgue measure, respectively the counting measure on A. Let us illustrate the Mellin transform by giving some examples of commonly considered distribution families.

Example 2.1.6 (Beta-Distribution):
We consider the family of Beta-Distributions $(\mathbb{P}_{a,b})_{(a,b)\in\mathbb{R}_+^2}$ with Lebesgue densities

$$f_{a,b}(x) := B(a,b)^{-1}x^{a-1}(1-x)^{b-1}\mathbb{1}_{(0,1)}(x)$$

for $x, a, b \in \mathbb{R}_+$. Here, B denotes the Beta-function $B(x,y) := \int_0^1 t^{x-1}(1-t)^{y-1}dt$ for $x, y \in \mathbb{C}$ with strictly positive real parts $\mathrm{Re}(x), \mathrm{Re}(y) > 0$. For all $a, b \in \mathbb{R}_+$ and $c \in (1-a, \infty)$ we have $\mathbb{P}_{a,b} \in \mathcal{P}_c^+$. The Mellin transform of the Beta-distribution is then given by

$$\mathcal{M}_c[\mathbb{P}_{a,b}](t) = B(a,b)^{-1}\int_{(0,1)} x^{c+a-2+it}(1-x)^{b-1}d\lambda(x) = \frac{B(c+a-1+it, b)}{B(a,b)}$$

for any $t \in \mathbb{R}$. If $b \in \mathbb{N}$, then the last expression simplifies to

$$\mathcal{M}_c[\mathbb{P}_{a,b}](t) = \prod_{j=1}^{b} \frac{a-1+j}{c+a-2+j+it} \qquad \text{for any } t \in \mathbb{R}.$$

Example 2.1.7 (Gamma-Distribution):
Let us consider the family of Gamma-Distributions $(\mathbb{P}_{p,\lambda})_{(p,\lambda)\in\mathbb{R}_+^2}$ with Lebesgue densities

$$f_{p,\lambda}(x) = \frac{\lambda^p}{\Gamma(p)}x^{p-1}\exp(-\lambda x)\mathbb{1}_{(0,\infty)}(x)$$

for $p, \lambda, x \in \mathbb{R}_+$. For all $p, \lambda \in \mathbb{R}_+$ and $c \in (1-p, \infty)$ we have $\mathbb{P}_{p,\lambda} \in \mathcal{P}_c^+$. The Mellin transform of the Gamma-Distribution is then given by

$$\mathcal{M}_c[\mathbb{P}_{p,\lambda}](t) = \lambda^{1-c-it}\frac{\Gamma(p+c-1+it)}{\Gamma(p)} \qquad \text{for any } t \in \mathbb{R}.$$

From the Definition 2.1.5 we can derive that $|\mathcal{M}_c[\mathbb{P}](t)| \leq \mathcal{M}_c[\mathbb{P}](0) = \mathbb{E}_{\mathbb{P}}[X^{c-1}]$. Furthermore, we can show the following property which is called the CONVOLUTION THEOREM for the Mellin transform.

Proposition 2.1.8 (Convolution theorem for the Mellin transform):
Let $c \in \mathbb{R}$ and $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_c^+$. Then $\mathbb{P} * \mathbb{Q} \in \mathcal{P}_c^+$ and

$$\mathcal{M}_c[\mathbb{P} * \mathbb{Q}] = \mathcal{M}_c[\mathbb{P}] \cdot \mathcal{M}_c[\mathbb{Q}].$$

Proof of Proposition 2.1.8. First we see that

$$\mathbb{E}_{\mathbb{P}*\mathbb{Q}}[X^{c-1}] = \int_{\mathbb{R}_+} x^{c-1}d(\mathbb{P}*\mathbb{Q})(x) = \int_{\mathbb{R}_+^2}(uv)^{c-1}d(\mathbb{P}\otimes\mathbb{Q})(u,v) = \mathbb{E}_{\mathbb{P}}[X^{c-1}]\mathbb{E}_{\mathbb{Q}}[X^{c-1}].$$

Analogously, we can show that for all $t \in \mathbb{R}$, $\mathcal{M}_c[\mathbb{P}*\mathbb{Q}](t) = \mathcal{M}_c[\mathbb{P}](t)\mathcal{M}_c[\mathbb{Q}](t)$. □

Now we will show that for two distribution having the same Mellin transform at a development point $c \in \mathbb{R}$ already implies that these distributions are identical.

Proposition 2.1.9 (Uniqueness of the Mellin transform)**:**
Let $c \in \mathbb{R}$ and $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_c^+$. If $\mathcal{M}_c[\mathbb{P}] = \mathcal{M}_c[\mathbb{Q}]$, then $\mathbb{P} = \mathbb{Q}$.

Proof of Proposition 2.1.9. It suffices to consider the case $c = 1$. Indeed, setting $\mu := \mathbb{E}_{\mathbb{P}}[X^{c-1}] = \mathcal{M}_c[\mathbb{P}](0) = \mathbb{E}_{\mathbb{Q}}[X^{c-1}]$ we define if $\mu \neq 0$ the distributions $\widetilde{\mathbb{P}}, \widetilde{\mathbb{Q}} \in \mathcal{P}_1^+$ through their Radon-Nikodym derivatives

$$\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}}(x) := \frac{x^{c-1}}{\mu}, \text{ respectively } \frac{d\widetilde{\mathbb{Q}}}{d\mathbb{Q}}(x) := \frac{x^{c-1}}{\mu}.$$

Then $\mu\mathcal{M}_1[\widetilde{\mathbb{P}}](t) = \mathcal{M}_c[\mathbb{P}](t) = \mathcal{M}_c[\mathbb{Q}](t) = \mu\mathcal{M}_1[\widetilde{\mathbb{Q}}](t)$ for any $t \in \mathbb{R}$. If $\mu = 0$ we already have $\mathbb{P} = \delta_{|\{0\}} = \mathbb{Q}$.
Let $c = 1$ and let $k \in \mathbb{R}_+$. Then for any $a < b \in \mathbb{R}_+$ we have

$$\int_{[-k,k]} \frac{a^{-it} - b^{-it}}{2\pi i t}\mathcal{M}_1[\mathbb{P}](t)d\lambda(t) = \int_{\mathbb{R}_+} \frac{1}{\pi} \int_{[-k,k]} \frac{\left(\frac{x}{a}\right)^{it} - \left(\frac{x}{b}\right)^{it}}{2it} d\lambda(t)d\mathbb{P}(x).$$

Considering the inner integral we see that by exploiting $\mu^{it} = \cos(\log(\mu)t) + i\sin(\log(\mu)t)$ for any $\mu \in \mathbb{R}_+$ we get

$$\begin{aligned}\int_{[-k,k]} \frac{\left(\frac{x}{a}\right)^{it} - \left(\frac{x}{b}\right)^{it}}{2it} d\lambda(t) &= \int_{[0,k]} \frac{\sin(\log(x/a)t) - \sin(\log(x/b)t)}{t} d\lambda(t) \\ &= g(k\log(x/a)) - g(k\log(x/b))\end{aligned}$$

for $g(t) := \frac{1}{\pi}\int_0^t \frac{\sin(u)}{u}du$. Now we have that $\lim_{t\to\infty} g(t) = 1/2$ and $\lim_{t\to-\infty} g(t) = -1/2$. This on the other-hand implies

$$g(k\log(x/a)) - g(k\log(x/b)) \to \begin{cases} 0 & , x \neq [a,b] \\ 1/2 & , x \in \{a,b\} \\ 1 & , x \in (a,b) \end{cases}.$$

Taking the limes inside the integral, which is allowed due to the fact that g is bounded and $\mathbb{P}$ is a probability measure, we get that

$$\lim_{k\to\infty} \int_{[-k,k]} \frac{a^{-it} - b^{-it}}{2\pi i t}\mathcal{M}_1[\mathbb{P}](t)d\lambda(t) = \mathbb{P}[(a,b)] + \frac{1}{2}(\mathbb{P}[\{a\}] + \mathbb{P}[\{b\}]).$$

Thus, we conclude from $\mathcal{M}_1[\mathbb{P}](t) = \mathcal{M}_1[\mathbb{Q}](t)$ that for any $a < b \in \mathbb{R}_+$ $\mathbb{Q}[(a,b)] + \frac{1}{2}(\mathbb{Q}[\{a\}] + \mathbb{Q}[\{b\}]) = \mathbb{P}[(a,b)] + \frac{1}{2}(\mathbb{P}[\{a\}] + \mathbb{P}[\{b\}])$. Now for any open interval $I = (a,b) \subset \mathbb{R}_+$ we can find a sequence $((a_n, b_n))_{n\in\mathbb{N}}$ with

- $a_n \leq a_{n+1} \leq a$, $b_n \geq b_{n+1} \geq b$ for all $n \in \mathbb{N}$;
- $\lim_{n\to\infty} a_n = a, \lim_{n\to\infty} b_n = b$ and
- $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are neither atoms of $\mathbb{P}$ nor atoms of $\mathbb{Q}$.

This is possible due to the fact that there are at most countable many atoms for each probability measure. Due to the continuity from above of the probability measure we deduce that $\mathbb{P}[I] = \mathbb{Q}[I]$ for any open interval $I \subset \mathbb{R}_+$. □

Now let us revisit Example 2.1.3 and Example 2.1.4 to show the convolution properties.

Example 2.1.10 (Example 2.1.3 continued)**:**
For all $a, \lambda \in \mathbb{R}_+$ and $c \in (-\infty, \lambda)$ the Mellin transform of the Log-Gamma distribution $\mathbb{P}_{a,\lambda}$ with density $f_{a,\lambda}(x) = \frac{\lambda^a}{\Gamma(a)} x^{-\lambda-1} \log(x)^{a-1} \mathbb{1}_{(1,\infty)}(x)$ is given by

$$\mathcal{M}_c[\mathbb{P}_{a,\lambda}](t) = \lambda^a(\lambda - c + 1 - it)^{-a}, \qquad \text{for all } t \in \mathbb{R}.$$

We deduce thus by application of the convolution theorem, Proposition 2.1.8, that $\mathcal{M}_c[\mathbb{P}_{a_1,\lambda} * \mathbb{P}_{a_2,\lambda}] = \mathcal{M}_c[\mathbb{P}_{(a_1+a_2,\lambda)}]$. Applying now the uniqueness of the Mellin transform, Proposition 2.1.9, we get $f_{a_1,\lambda} * f_{a_2,\lambda} = f_{a_1+a_2,\lambda}$.

Example 2.1.11 (Example 2.1.4 continued)**:**
For all $\mu \in \mathbb{R}$, $\lambda \in \mathbb{R}_+$ and $c \in \mathbb{R}$ the Mellin transform of the Log-Normal distribution $(\mathbb{P}_{\mu,\lambda^2})$ with density $f_{\mu,\lambda^2}(x) = \frac{1}{\sqrt{2\pi}\lambda x} \exp(-(\log(x) - \mu)^2/2\lambda^2) \mathbb{1}_{\mathbb{R}_+}(x)$ is given by

$$\mathcal{M}_c[\mathbb{P}_{\mu,\lambda}](t) = \exp(\mu(c - 1 + it)) \exp(\lambda^2 \frac{(c-1+it)^2}{2}), \qquad \text{for all } t \in \mathbb{R}.$$

We deduce thus by application of the convolution theorem, Proposition 2.1.8, that $\mathcal{M}_c[\mathbb{P}_{a_1,\lambda_1^2} * \mathbb{P}_{a_2,\lambda_2^2}] = \mathcal{M}_c[\mathbb{P}_{(a_1+a_2,\lambda_1^2+\lambda_2^2)}]$. Applying now the uniqueness of the Mellin transform, Proposition 2.1.9, we get that $f_{a_1,\lambda_1^2} * f_{a_2,\lambda_2^2} = f_{a_1+a_2,\lambda_1^2+\lambda_2^2}$.

2.2 The Mellin transform for integrable functions

In the last section we introduced the Mellin transform for arbitrary probability measures $\mathbb{P}$ on $(\mathbb{R}_+, \mathscr{B})$. For probability measures dominated by the Lebesgue measure, we have seen that the Mellin transform can be expressed using the Radon-Nikodym derivative of $\mathbb{P}$ with respect to the Lebesgue measure. For the sake of simplicity we will call any Radon-Nikodym derivative with respect to the Lebesgue measure from now on a density with respect to the Lebesgue measure or short a density.

In this section, we will build a theory of the Mellin transform from a more functional analytical point of view. To do so, we construct the Mellin transform and analyze its properties similar to the theory of Fourier transform. A well-structured introduction to the theory of Fourier transform can be found for instance in Werner (2006). The introduction of Werner (2006) inspired the presented theory.

Let us now introduce some frequently used notation and definitions from the theory of $\mathbb{L}^p$-spaces. More precisely, we consider for a Borel-measurable weight function $\omega : \mathbb{R} \to \mathbb{R}_+$ the corresponding weighted semi-norm by $\|h\|_\omega := \int_{\mathbb{R}_+} |h(x)|\omega(x)d\lambda(x)$ for a measurable function $h : \mathbb{R}_+ \to \mathbb{C}$. Denote by $L^1(\mathbb{R}_+, \omega)$ the set of all measurable functions with finite $\| \,.\, \|_{L^1(\mathbb{R}_+,\omega)}$-semi-norm. For $h_1, h_2 \in L^1(\mathbb{R}_+, \omega)$ we can define the equivalence relation $\sim$ by $h_1 \sim h_2 :\Leftrightarrow h_1 = h_2$ Lebesgue-almost everywhere. As usual, we denote by $\mathbb{L}^1(\mathbb{R}_+, \omega) := \{[h] : h \in L^1(\mathbb{R}_+, \omega)\}$ the set of all equivalence classes $[h]$ with respect to $\sim$. On this set, we can define the norm $\|[h]\|_{\mathbb{L}^1(\mathbb{R}_+,\omega)} := \|h\|_{L^1(\mathbb{R}_+,\omega)}$ where $h \in L^1(\mathbb{R}_+, \omega)$ is a representant of $[h] \in \mathbb{L}^1(\mathbb{R}_+, \omega)$. Since the right-hand-side of the definition is independent of the choice of the representant $h \in [h]$ the norm is well-defined and in fact a norm. By abuse of notation, we will write $\|h\|_{\mathbb{L}^1(\mathbb{R}_+,\omega)}$ instead of $\|[h]\|_{\mathbb{L}^1(\mathbb{R}_+,\omega)}$ and $h \in \mathbb{L}^1(\mathbb{R}_+, \omega)$ instead of $[h] \in \mathbb{L}^1(\mathbb{R}_+, \omega)$ keeping in mind, that h always stands for the whole equivalent class which h is a member of.

2.2.1 Multiplicative convolution

The following Lemma ensures that we can define the multiplicative convolution of two functions $h_1, h_2 \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$, where the weight function x^{c-1} for $c \in \mathbb{R}$ is defined by $x^{c-1} : \mathbb{R}_+ \to \mathbb{R}_+, y \mapsto y^{c-1}$.

Lemma 2.2.1:
Let $c \in \mathbb{R}$ and $h_1, h_2 \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$. Then for any $y \in \mathbb{R}_+$ the function $I_y : \mathbb{R}_+ \mapsto \mathbb{R}, x \mapsto I_y(x) := h_1(y/x)h_2(x)x^{-1}$ is measurable and integrable for almost all $y \in \mathbb{R}_+$. Furthermore, the function $h_1 * h_2 : \mathbb{R}_+ \to \mathbb{R}, y \mapsto \int_{\mathbb{R}_+} I_y(x)d\lambda(x)$ is weighted integrable, that is $h_1 * h_2 \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ with

$$\|h_1 * h_2\|_{\mathbb{L}^1(\mathbb{R}_+,x^{c-1})} \leq \|h_1\|_{\mathbb{L}^1(\mathbb{R}_+,x^{c-1})}\|h_2\|_{\mathbb{L}^1(\mathbb{R}_+,x^{c-1})}.$$

Proof of Lemma 2.2.1. The measurability of I_y is obvious. The fact, that I_y is integrable for almost all $y \in \mathbb{R}_+$ is equivalent to the case that the set $N := \{y \in \mathbb{R}_+ : \int_{\mathbb{R}_+} |h_1(y/x)||h_2(x)|x^{-1}d\lambda(x) = \infty\}$ is a Lebesgue-null set. To show this, let us define $\widetilde{h}_1 := x^{c-1}|h_1|\|h_1\|^{-1}_{\mathbb{L}^1(\mathbb{R}_+,x^{c-1})}$ and analogously $\widetilde{h}_2$. Then both, $\widetilde{h}_1, \widetilde{h}_2$ are densities. Denoting $\mathbb{P}_1, \mathbb{P}_2$ the corresponding distributions on

$(\mathbb{R}_+, \mathscr{B}_+)$ we see that $\mathbb{P}_1 * \mathbb{P}_2$ has the density

$$(\widetilde{h}_1 * \widetilde{h}_2)(y) = \frac{y^{c-1}}{\|h_1\|_{\mathbb{L}^1(\mathbb{R}_+, x^{c-1})}\|h_2\|_{\mathbb{L}^1(\mathbb{R}_+, x^{c-1})}} \int_{\mathbb{R}_+} |h_1(y/x)||h_2(x)|x^{-1} d\lambda(x) \tag{2.2}$$

Since $(\widetilde{h}_1 * \widetilde{h}_2)$ is a density, the set N needs to be a Lebesgue-null set and $\|\widetilde{h}_1 * \widetilde{h}_2\|_{\mathbb{L}^1(\mathbb{R}_+, x^0)} = 1$. On the other hand, we can directly derive from (2.2) that

$$\begin{aligned} \|h_1 * h_2\|_{\mathbb{L}^1(\mathbb{R}_+, x^{c-1})} &\leq \int_{\mathbb{R}_+} y^{c-1} \int_{\mathbb{R}_+} |h_1(y/x)||h_2(x)|x^{-1} d\lambda(x) d\lambda(y) \\ &= \|h_1\|_{\mathbb{L}^1(\mathbb{R}_+, x^{c-1})}\|h_2\|_{\mathbb{L}^1(\mathbb{R}_+, x^{c-1})}. \end{aligned}$$

□

It is worth stressing out, that the function $h_1 * h_2$ is not dependent on the choice of $c \in \mathbb{R}$. Now we are able to define the multiplicative convolution for $\mathbb{L}^1(\mathbb{R}_+, x^{c-1})$-functions.

Definition 2.2.2 (Multiplicative convolution of $\mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ functions)**:**
Let $c \in \mathbb{R}$ and $h_1, h_2 \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$. Then we call the function $h_1 * h_2 \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ with

$$(h_1 * h_2)(y) := \int_{\mathbb{R}_+} h_1(y/x)h_2(x)x^{-1} d\lambda(x) \quad \text{for any } y \in \mathbb{R}_+,$$

the MULTIPLICATIVE CONVOLUTION of h_1 and h_2.

By definition of the multiplicative convolution operator $* : \mathbb{L}^1(\mathbb{R}_+, x^{c-1}) \times \mathbb{L}^1(\mathbb{R}_+, x^{c-1}) \to \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ with $(h_1, h_2) \mapsto h_1 * h_2$ is associative and commutative. Further, for two probability measures $\mathbb{P}_1, \mathbb{P}_2$ with densities h_1, h_2 the multiplicative convolution $h_1 * h_2$ of h_1 and h_2 coincides with the density of the multiplicative convolution $\mathbb{P}_1 * \mathbb{P}_2$ of the measures $\mathbb{P}_1$ and $\mathbb{P}_2$. We will now define the Mellin transform for a function $h \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$.

2.2.2 The Mellin transform

For $h \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ the function $x \mapsto x^{c-1+it}h(x)$ is absolutely integrable for any $t \in \mathbb{R}$ which allows us to define the Mellin transform of h in the following way.

Definition 2.2.3 (The Mellin transform for $\mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ functions)**:**
Let $c \in \mathbb{R}$ and $h \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$. Then we define the MELLIN TRANSFORM $\mathcal{M}_c[h]$ of the function h developed in the point c as the function $\mathcal{M}_c[h] : \mathbb{R} \to \mathbb{C}$, where

$$\mathcal{M}_c[h](t) := \int_{\mathbb{R}_+} x^{c-1+it}h(x) d\lambda(x) \text{ for any } t \in \mathbb{R}.$$

We observe for $\mathbb{P} \in \mathcal{P}_c^+$ with Lebesgue density f, that $\mathcal{M}_c[\mathbb{P}] = \mathcal{M}_c[f]$, which motivates the abuse of notation. From the definition of the Mellin transform we can derive the following properties.

Proposition 2.2.4 (Calculation rules for Mellin transforms):
Let $c, p \in \mathbb{R}$ and $\kappa \in \mathbb{R}_+$.

(i) If $h \in \mathbb{L}^1(\mathbb{R}_+, x^{c+p-1})$ then $\mathcal{M}_c[x^p h] = \mathcal{M}_{c+p}[h]$

(ii) If $h \in \mathbb{L}^1(\mathbb{R}_+, x^{c/\kappa-1})$ then $\mathcal{M}_c[h \circ x^\kappa](t) = \kappa^{-1}\mathcal{M}_{c/\kappa}[h](t/\kappa)$ for any $t \in \mathbb{R}$.

(iii) If $h \in \mathbb{L}^1(\mathbb{R}_+, x^{-c-1})$ then $\mathcal{M}_c[h(x^{-1})](t) = \mathcal{M}_{-c}[h](-t)$ for any $t \in \mathbb{R}$.

(iv) If $h \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ then $\mathcal{M}_c[h(\kappa x)](t) = \kappa^{-c-it}\mathcal{M}_c[h](t)$ for any $t \in \mathbb{R}$.

(v) If $h \in \mathbb{L}^1(\mathbb{R}_+, x^c)$, with $c > 0$, and $S_h(y) := \int_{(y,\infty)} h(x)d\lambda(x), y \in \mathbb{R}_+$, then $\mathcal{M}_c[S_h](t) = (c+it)^{-1}\mathcal{M}_{c+1}[h](t)$ for any $t \in \mathbb{R}$.

Proof of Proposition 2.2.4. The claim (i) is obvious. For (ii) we see that

$$\begin{aligned}\mathcal{M}_c[h(x^\kappa)](t) &= \kappa^{-1}\int_{\mathbb{R}_+} x^{c-\kappa+it}\kappa x^{\kappa-1}h(x^\kappa)d\lambda(x)\\ &= \kappa^{-1}\int_{\mathbb{R}_+} x^{c/\kappa-1+it/\kappa}h(x)d\lambda(x) = \kappa^{-1}\mathcal{M}_{c/\kappa}[h](t/\kappa)\end{aligned}$$

for any $t \in \mathbb{R}$ by a change of variables. Analogously, we can show (iii) and (iv) by a change of variables. For (v) we see that for any $t \in \mathbb{R}$ by application of Fubini–Tonelli,

$$\begin{aligned}\mathcal{M}_c[S_h](t) &= \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} x^{c-1+it}\mathbb{1}_{(0,y)}(x)h(y)d\lambda(y)d\lambda(x) = \int_{\mathbb{R}_+} h(y)\int_{(0,y)} x^{c-1+it}d\lambda(x)d\lambda(y)\\ &= (c+it)^{-1}\int_{\mathbb{R}_+} y^{c+it}h(y)d\lambda(y) = (c+it)\mathcal{M}_{c+1}[h](t).\end{aligned}$$

□

We will show that the convolution theorem, Proposition 2.1.8, holds also true on $\mathbb{L}^1(\mathbb{R}_+, x^{c-1})$.

Proposition 2.2.5 (Convolution theorem for $\mathbb{L}^1(x^{c-1}, \mathbb{R}_+)$ functions):
Let $h_1, h_2 \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ for a $c \in \mathbb{R}$. Then

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

Proof of Proposition 2.2.5. In the proof of Lemma 2.2.1 we have already seen that the function $y \mapsto \int_0^\infty |h_1(y/x)||h_2(x)|x^{-1}dx$ belongs to $\mathbb{L}^1(\mathbb{R}_+, x^{c-1})$. This allows for the following usage of the Fubini–Tonelli theorem

$$\begin{aligned}\mathcal{M}_c[h_1 * h_2](t) &= \int_{\mathbb{R}_+} y^{c-1+it}h_1 * h_2(y)d\lambda(y)\\ &= \int_{(0,\infty)} h_2(x)x^{-1}\int_{(0,\infty)} y^{c-1+it}h_1(y/x)d\lambda(y)d\lambda(x)\\ &= \int_{(0,\infty)} h_2(x)x^{c-1+it}\int_{(0,\infty)} y^{c-1+it}h_1(y)d\lambda(y)d\lambda(x) = \mathcal{M}_c[h_1](t)\cdot\mathcal{M}_c[h_2](t)\end{aligned}$$

for any $t \in \mathbb{R}$. □

2.3 The Mellin transform for square-integrable functions

In this section we expand the Mellin transform to square-integrable functions and derive the frequently used properties such as the inversion formula, a Plancherel identity, and develop a theory of Sobolev spaces for the Mellin transform. An equivalent introduction to the Fourier transform can be found in Werner (2006) which inspired the upcoming section.
Let us, similar to $\mathbb{L}^1(\mathbb{R}_+, x^{c-1})$, define

$$\mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) = \{h : \mathbb{R}_+ \to \mathbb{C} : \|h\|_{x^{2c-1}}^2 := \int_{\mathbb{R}_+} |h(x)|^2 x^{2c-1} d\lambda(x) < \infty\}/\sim,$$

the quotient set of all measurable functions $h : \mathbb{R}_+ \to \mathbb{C}$ with finite $\mathbb{L}^2(\mathbb{R}, x^{2c-1})$ norm, where $\sim$ is the equivalence relation defined in the last section. We endow the set $\mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ with the induced inner product $\langle h_1, h_2\rangle_{x^{2c-1}} := \int_{\mathbb{R}_+} h_1(x)\overline{h_2(x)} x^{2c-1} d\lambda(x)$ for $h_1, h_2 \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$. Analogously, we define $\mathbb{L}^2(\mathbb{R})$, respectively $\mathbb{L}^1(\mathbb{R})$, the set of all measurable functions $H : \mathbb{R} \to \mathbb{C}$ with finite $\mathbb{L}^2(\mathbb{R})$-norm, that is $\|H\|_{\mathbb{R}}^2 := \int_{\mathbb{R}} |H(x)|^2 d\lambda(x) < \infty$, and endow it with the inner product $\langle H_1, H_2\rangle_{\mathbb{R}} := \int_{\mathbb{R}} H_1(x)\overline{H_2(x)} d\lambda(x)$. Furthermore, let $\mathbb{L}^1(\mathbb{R})$ be the quotient set of all integrable, complex-valued measurable function with norm $\|H\|_{\mathbb{L}^1(\mathbb{R})} := \int_{\mathbb{R}} |H(x)| d\lambda(x)$ for $H \in \mathbb{L}^1(\mathbb{R})$.

2.3.1 Multiplicative convolution

Lemma 2.3.1:
Let $h_1, h_2 \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ and $h_2 \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ for a $c \in \mathbb{R}$. Then $h_1 * h_2 \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1}) \cap \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ with $\|h_1 * h_2\|_{x^{2c-1}} \leq \|h_1\|_{\mathbb{L}^1(\mathbb{R}_+, x^{c-1})} \|h_2\|_{x^{2c-1}}$.

Proof of Lemma 2.3.1. Due to Lemma 2.2.1 we have $h_1 * h_2 \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$. It remains to show that $h_1 * h_2 \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$. By the application of the generalized Minkowski inequality, we see

$$\begin{aligned}
\|h_1 * h_2\|_{x^{2c-1}}^2 &= \int_{\mathbb{R}_+} \left| \int_{\mathbb{R}_+} h_1(y/x) h_2(x) x^{-1} y^{c-1/2} d\lambda(x) \right|^2 d\lambda(y) \\
&\leq \left(\int_{\mathbb{R}_+} \left(\int_{\mathbb{R}_+} |h_1(y/x)|^2 |h_2(x)|^2 x^{-2} y^{2c-1} d\lambda(y) \right)^{1/2} d\lambda(x) \right)^2 \\
&= \|h_1\|_{x^{2c-1}}^2 \|h_2\|_{\mathbb{L}^1(\mathbb{R}_+, x^{c-1})}^2.
\end{aligned}$$

□

2.3.2 The Mellin transform

For $c \in \mathbb{R}$ let us first revisit the definition of the Mellin transform on $\mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ functions. Secondly, we deduce a notion of a Mellin transform on $\mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$. Let $h \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$. Then for any $t \in \mathbb{R}$, by a change of variable, $\int_0^\infty x^{c-1+it} h(x) dx = \int_{-\infty}^\infty \exp(-itx) \exp(-xc) h(\exp(-x)) dx$ which implies

$$\mathcal{M}_c[h](t) = (2\pi)^{1/2} \mathcal{F}[H](t), \tag{2.3}$$

where $H(x) := \exp(-xc) h(\exp(-x))$ and $\mathcal{F}[H] : \mathbb{R} \to \mathbb{C}, t \mapsto (2\pi)^{-1/2} \int_{\mathbb{R}} \exp(-itx) H(x) d\lambda(x)$ is the Fourier transform for $H \in \mathbb{L}^1(\mathbb{R})$. A definition of the Fourier transform and some of its major properties are collected in Section 2.6. The identity in equation (2.3) motivates the following construction. Let us define the diffeomorphism $\varphi : \mathbb{R} \to \mathbb{R}_+, x \mapsto \exp(-x)$. Furthermore, let us denote by $\mathrm{M}(\mathbb{R})$, respectively $\mathrm{M}(\mathbb{R}_+)$ the set of all measurable, complex-valued functions on $\mathbb{R}$, respectively on $\mathbb{R}_+$. Then the operator

$$\Phi_c : \mathrm{M}(\mathbb{R}_+) \to \mathrm{M}(\mathbb{R}), h \mapsto \Phi(h) := \varphi^c \cdot (h \circ \varphi)$$

is isomorphic with inverse $\Phi_c^\dagger : \mathrm{M}(\mathbb{R}) \to \mathrm{M}(\mathbb{R}_+).H \mapsto x^{-c} \cdot (H \circ \varphi^\dagger)$, where $\varphi^\dagger(x) := -\log(x), x \in \mathbb{R}_+$. Since φ is a diffeomorphism we get for $h_1 \sim h_2$, $h_1, h_2 \in \mathrm{M}(\mathbb{R}_+)$ that $\Phi_c[h_1] \sim \Phi_c[h_2]$. Therefore we can interpret the operator Φ_c as an operator $\Phi_c : \mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) \to \mathrm{M}(\mathbb{R})/\sim$. More precisely, we show the following mapping property.

Lemma 2.3.2:
The mapping $\Phi_c : \mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) \to \mathbb{L}^2(\mathbb{R})$ is an isomorphism with inverse $\Phi_c^\dagger$.

Proof of Lemma 2.3.2. Let $h \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$. Then,

$$\begin{aligned}\int_{\mathbb{R}} |\Phi_c[h](x)|^2 d\lambda(x) &= \int_{\mathbb{R}} \exp(-2xc)|h(\exp(-x))|^2 d\lambda(x) \\ &= \int_{\mathbb{R}_+} x^{2c-1}|h(x)|^2 d\lambda(x) < \infty,\end{aligned}$$

which implies$\Phi_c[\mathbb{L}^2(\mathbb{R}, x^{-1})] \subseteq \mathbb{L}^2(\mathbb{R})$. Analogously, we see that $\Phi_c^\dagger[\mathbb{L}^2(\mathbb{R})] \subseteq \mathbb{L}^2(\mathbb{R}_+, x^{-1})$. □

Let us now define the operator

$$\mathcal{F}_2 : \mathbb{L}^2(\mathbb{R}) \to \mathbb{L}^2(\mathbb{R}), H \mapsto \lim_{k\to\infty} \mathcal{F}[H\mathbb{1}_{[-k,k]}],$$

the so-called FOURIER-PLANCHEREL transform, where the limit is understood in the sense of $\mathbb{L}^2(\mathbb{R})$ convergences, compare Theorem 2.6.3. As an abuse of notation we will drop the index of the Fourier-Plancherel transform. Now, Equation (2.3) motivates the following Definition.

Definition 2.3.3 (The Mellin transform for $\mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ functions)**:**
For $c \in \mathbb{R}$ and $h \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ we define the MELLIN TRANSFORM $\mathcal{M}_c[h]$ of h developed in c as the function $\mathcal{M}_c[h] : \mathbb{R} \to \mathbb{C}$ where

$$\mathcal{M}_c[h] := (2\pi)^{1/2}(\mathcal{F} \circ \Phi_c)(h).$$

If $h \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) \cap \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$, then the two notions of the Mellin transforms, compare Definition 2.2.3 and Definition 2.3.3, are identical. This property is captured in the next Proposition.

Proposition 2.3.4:
Let $c \in \mathbb{R}$ and $h \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) \cap \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$. Then for any $t \in \mathbb{R}$,

$$(2\pi)^{1/2}(\mathcal{F} \circ \Phi_c)(t) = \int_{\mathbb{R}_+} x^{c-1+it} h(x) d\lambda(x).$$

Proof of Proposition 2.3.4. For $h \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) \cap \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ we have $\Phi_c[h] \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. In this situation, the Fourier and the Fourier-Plancherel transformation coincides, see Proposition 2.6.5. From this we deriveby a change of variable that for any $t \in \mathbb{R}$,

$$(2\pi)^{1/2}\mathcal{F}[\Phi_c[h]](t) = \int_{\mathbb{R}} \exp(-itx)\Phi_c[h](x)d\lambda(x) = \int_{\mathbb{R}_+} x^{c-1+it}h(x)d\lambda(x).$$

□

We will now derive important properties of the Mellin transform which will be used frequently in this dissertation.

Proposition 2.3.5 (Plancherel identity for Mellin transforms)**:**
Let $c \in \mathbb{R}$. Then for all $h_1, h_2 \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$, the Plancherel identity holds true, that is

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

Furthermore, the Parseval identity holds true, that is $\|h_1\|^2_{x^{2c-1}} = (2\pi)^{-1}\|\mathcal{M}_c[h_1]\|^2_{\mathbb{R}}$.

Proof of Proposition 2.3.5. For $h_1, h_2 \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ we have by a change of variable

$$\begin{aligned}\langle h_1, h_2 \rangle_{x^{2c-1}} &= \langle \Phi_c[h_1], \Phi_c[h_2]\rangle_{\mathbb{R}} \\ &= (2\pi)^{-1}\langle \mathcal{F}[\Phi_c[h_1]], \mathcal{F}[\Phi_c[h_2]]\rangle_{\mathbb{R}}\end{aligned}$$

and an application of the Plancherel identity, Proposition 2.6.4, for the Fourier-Plancherel transform. This shows the first claim. The Parseval identity follows from the Plancherel identity and vice versa. □

As a composition of isomorphisms $\mathcal{M}_c : \mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) \to \mathbb{L}^2(\mathbb{R})$ is an isomorphism. Let us denote by $\mathcal{M}_c^\dagger : \mathbb{L}^2(\mathbb{R}) \to \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ its inverse. Then we can see that $\mathcal{M}_c^\dagger = (2\pi)^{1/2}\Phi_c^\dagger \circ \mathcal{F}^\dagger$ where $\mathcal{F}^\dagger$ is the inverse of the Fourier-Plancherel transform. We can therefore derive the following explicit representation of the inverse Mellin transform for a special case. More precisely, in the case of $H \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ we can express a representant of $\mathcal{M}_c^\dagger[H]$ explicitly through the following formula.

Proposition 2.3.6 (The inverse Mellin transform)**:**
For $H \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ the inverse Mellin transform $\mathcal{M}_c^\dagger : \mathbb{L}^2(\mathbb{R}) \to \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ for $c \in \mathbb{R}$ of H is explicitly given by

$$\mathcal{M}_c^\dagger[H](x) = \frac{1}{2\pi}\int_{\mathbb{R}} x^{-c-it}H(t)d\lambda(t) \quad \text{for any } x \in \mathbb{R}_+.$$

Proof of Proposition 2.3.6. For $H \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ we can express the inverse of the Fourier-Plancherel transform by $\mathcal{F}^\dagger[H](x) = (2\pi)^{-1/2}\int_{\mathbb{R}} \exp(itx)H(t)d\lambda(t)$, compare Theorem 2.6.6. On the other hand for any $H' \in \mathbb{L}^2(\mathbb{R})$ we have $\Phi_c^\dagger[H'](x) = x^{-c}H'(-\log(x))$ for any $x \in \mathbb{R}_+$. This implies

$$\mathcal{M}_c^\dagger[H](x) = \frac{1}{2\pi}\Phi_c^{-1}[y \mapsto \int_{\mathbb{R}} \exp(ity)H(t)d\lambda(t)](x) = \frac{1}{2\pi}\int_{\mathbb{R}} x^{-c-it}H(t)d\lambda(t)$$

for any $x \in \mathbb{R}_+$. □

For $H_1, H_2 \in \mathbb{L}^1(\mathbb{R})$ we denote the additive convolution $H_1 *_+ H_2 \in \mathbb{L}^1(\mathbb{R})$ as usual by

$$H_1 *_+ H_2(y) := \int_{\mathbb{R}} H_1(y-x)H_2(x)d\lambda(x) \quad y \in \mathbb{R},$$

compare Definition 2.6.7.

Proposition 2.3.7 (Product rule):
Let $h_1, h_2 \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ such that $\mathcal{M}_c[h_1], \mathcal{M}_c[h_2] \in \mathbb{L}^1(\mathbb{R})$. Then $h_1h_2 \in \mathbb{L}^1(\mathbb{R}_+, x^{2c-1})$ and

$$\frac{1}{2\pi}\mathcal{M}_{2c}[h_1h_2] = \mathcal{M}_c[h_1] *_+ \mathcal{M}_c[h_2].$$

Proof of Proposition 2.3.7. First, the Cauchy-Schwarz inequality implies $h_1h_2 \in \mathbb{L}^1(\mathbb{R}, x^{2c-1})$, and hence $\mathcal{M}_{2c}[h_1h_2]$ is well-defined with $\mathcal{M}_{2c}[h_1h_2] = (2\pi)^{1/2}\mathcal{F}[\Phi_{2c}[h_1h_2]]$ where $\mathcal{F}$ denotes the Fourier transform for $\mathbb{L}^1(\mathbb{R})$ functions. Since $\mathcal{M}_c[h_1], \mathcal{M}_c[h_2] \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ we have $\mathcal{M}_c[h_1] *_+ \mathcal{M}_c[h_2] \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. Due to the convolution theorem of the Fourier transform we have $\mathcal{F}[(\mathcal{M}_c[h_1] *_+ \mathcal{M}_c[h_2])](t) = (2\pi)^{1/2}\mathcal{F}[\mathcal{M}_c[h_1]](t)\mathcal{F}[\mathcal{M}_c[h_2]](t)$, for $t \in \mathbb{R}$, from which we deduce

$$\mathcal{F}[(\mathcal{M}_c[h_1] *_+ \mathcal{M}_c[h_2])](t) = (2\pi)^{3/2}\Phi_c[h_1](-t)\Phi_c[h_1](-t) = (2\pi)^{3/2}\Phi_{2c}[h_1h_2](-t).$$

Now applying the Fourier transform on both sides implies the statement. □

Example 2.3.8 (The lsi function):
Let us define the logarithmic si function as

$$\text{lsi} : \mathbb{R}_+ \to \mathbb{R}, \quad x \mapsto \text{lsi}(x) := \frac{\sin(\log(x))}{\pi \log(x)}.$$

Then we can see that $\text{lsi} \in \mathbb{L}^2(\mathbb{R}, x^{-1})$ with

$$\mathcal{M}_0[\text{lsi}](t) = \mathbb{1}_{[-1,1]}(t), \quad t \in \mathbb{R},$$

and thus $\mathcal{M}_0[\text{lsi}] \in \mathbb{L}^1(\mathbb{R})$. In other words, the Mellin transform of the lsi function is a scaled rectangle function. We deduce using Proposition 2.3.7 that $\text{lsi}^2 \in \mathbb{L}^1(\mathbb{R}, x^{-1})$, where

$$\mathcal{M}_0[\text{lsi}^2](t) = (\mathbb{1}_{[-1,1]} *_+ \mathbb{1}_{[-1,1]})(t) = (2 - |t|)\mathbb{1}_{[-2,2]}(t),$$

for $t \in \mathbb{R}$. Here the Mellin transform of lsi^2 is a scaled triangle function.

2.3.3 Mellin Sobolev spaces

In analogy to the theory of Sobolev spaces for Fourier transforms we define a Sobolev space for the Mellin transform through the decay of the Mellin transform. After that, we will analyze what degree of smoothness characterization is given through these spaces.

Definition 2.3.9 (The Mellin Sobolev-space):
For $c \in \mathbb{R}$ and $s \in \mathbb{R}_+$ we define the MELLIN-SOBOLEV space through

$$\mathbb{W}^s_c(\mathbb{R}_+) := \{h \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) : \|(1+|t|^2)^{s/2}\mathcal{M}_c[h]\|^2_{\mathbb{R}} < \infty\}.$$

By the definition of the Mellin-Sobolev spaces, we see that $\Phi_c[\mathbb{W}_c^s(\mathbb{R}_+)] = \{H \in \mathbb{L}^2(\mathbb{R}) : \|(1+|t|^2)^{s/2}\mathcal{F}[H]\|_{\mathbb{R}}^2 < \infty\} = W^s(\mathbb{R})$ which is the usual Fourier-Sobolev space, see Equation (2.5). For integer-valued $s \in \mathbb{R}_+$, that is $s \in \mathbb{N}$, this coincides with the Sobolev spaces characterized via the existence of weak derivatives, see Definition 2.6.10. Through this consideration we get the following version of the Lemma of Sobolev for Mellin transforms. Here, in contrary to the rest of this thesis, we strongly distinguish between the equivalence class $[h]$ and a representant h using the bold brackets, that is $[\boldsymbol{h}]$.

Theorem 2.3.10 (Lemma of Sobolev)**:**
Let $m, k \in \mathbb{N}$ such that $m > k + 1/2$. Then for each equivalence class $[\boldsymbol{h}] \in \mathbb{W}_c^m(\mathbb{R}_+)$ exists a representant of $[\boldsymbol{h}]$ which is k-times continuously differentiable.

Proof of Theorem 2.3.10. Let $h \in \mathbb{W}_c^m(\mathbb{R}_+)$. If for the equivalence class $[\widetilde{\boldsymbol{h}}]$ with $\widetilde{h} : \mathbb{R}_+ \to \mathbb{C}, x \mapsto x^c h$ exists a k-time differentiable representant , so does for the equivalence class $[\boldsymbol{h}]$. Therefore, we restrict ourselves to the case that $c = 0$. For $[\boldsymbol{h}] \in \mathbb{W}_0^m(\mathbb{R}_+)$ we have $\Phi_0\,[[\boldsymbol{h}]] = [\boldsymbol{\Phi_0}[\boldsymbol{h}]] \in \{H \in \mathbb{L}^2(\mathbb{R}) : \|(1+|t|)^{m/2}\mathcal{F}[H]\|^2 < \infty\}$. Applying the Lemma of Sobolev, Theorem 2.6.11, for the Fourier transform, we get that there exists a representant $\overline{H} \in [\boldsymbol{\Phi_0}[\boldsymbol{h}]]$ that is k-times continuously differentiable. Therefore, $\Phi_0^\dagger[\overline{H}]$ is k-times continuously differentiable as a composition of such functions. Furthermore, since $\Phi_0^\dagger$ is a diffeomorphism, we get that $\Phi_0^\dagger[\overline{H}]$ is representant of $[\boldsymbol{h}]$ which shows the claim. □

Our aim is now to define a weak derivative operator for the Mellin transform which characterizes the Mellin Sobolev spaces for $s \in \mathbb{N}$ in the same manner as the usual weak derivative operator does for the Fourier-Sobolev spaces. To do so, let us define the set $C_0^\infty(\mathbb{R}_+)$ as the set of all smooth, complex-valued functions with compact support in $\mathbb{R}_+$, that is for any $\beta \in \mathbb{N}$ and any $h \in C_0^\infty(\mathbb{R}_+)$ the derivatives

$$\mathrm{D}^\beta[h] := \frac{d^\beta}{dx^\beta}h$$

exist and vanish for $x \to 0$ and $x \to \infty$. Through the natural inclusion of $C_0^\infty(\mathbb{R}_+)$ in $\mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ for any $c \in \mathbb{R}$ we can show the following property.

Proposition 2.3.11:
For any $\beta \in \mathbb{N}$, $c \in \mathbb{R}$ and $h \in C_0^\infty(\mathbb{R}_+)$,

$$\mathcal{M}_c[\mathrm{D}^\beta[h]](t) = (-1)^\beta \frac{\Gamma(c+it)}{\Gamma(c-\beta+it)}\mathcal{M}_{c-\beta}[h](t)$$

for any $t \in \mathbb{R}$.

Proof of Proposition 2.3.11. First we see for $\beta = 1$ that

$$\begin{aligned}\mathcal{M}_c[\mathrm{D}[h]](t) &= \int_{\mathbb{R}_+} \mathrm{D}[h](x)x^{c-1+it}d\lambda(x) \\ &= \left[h(x)x^{c-1+it}\right]_0^\infty - \int_{\mathbb{R}_+}(c-1+it)x^{c-2+it}h(x)d\lambda(x) \\ &= -(c-1+it)\mathcal{M}_{c-1}[h](t).\end{aligned}$$

Now the claim follows by induction and $\frac{\Gamma(c+it)}{\Gamma(c-\beta+it)} = \prod_{j=1}^\beta (c-j+it)$ for any $t \in \mathbb{R}$. □

Thus, the usual differential operator does not interact with the Mellin transform as it does with the Fourier transform, compare Lemma 2.6.12 and the previous discussion there. A more natural operator can be defined in the following way. We set

$$\mathrm{S} : C_0^\infty(\mathbb{R}_+) \to C_0^\infty(\mathbb{R}_+), h \mapsto (x \mapsto -x\mathrm{D}^1[h](x))$$

and for $\beta \in \mathbb{N}$ define $\mathrm{S}^\beta := \mathrm{S}^{\beta-1} \circ \mathrm{S}$ where $\mathrm{S}^1 := \mathrm{S}$. Then we can show the following property.

Proposition 2.3.12:
For $\beta \in \mathbb{N}$, $c \in \mathbb{R}$ and $h \in C_0^\infty(\mathbb{R}_+)$,

$$\mathcal{M}_c[\mathrm{S}^\beta \circ h](t) = (c + it)^\beta \mathcal{M}_c[h](t)$$

for any $t \in \mathbb{R}$.

Proof of Proposition 2.3.12. First for $\beta = 1$ we have

$$\mathcal{M}_c[\mathrm{S}[h]](t) = -\mathcal{M}_{c+1}[\mathrm{D}^1[h]](t) = (c + it)\mathcal{M}_c[h]$$

by application of Proposition 2.3.11 for any $t \in \mathbb{R}$. The claim follows then by an induction. □

The operator S^β can also be motivated by the following consideration for $c = 0$. For $h \in C_0^\infty(\mathbb{R}_+)$ we see that $\Phi_0[h] \in C_0^\infty(\mathbb{R})$. Applying, the operator D^1 we obtain

$$\mathrm{D}^1[\Phi_0[h]] = -\varphi(\mathrm{D}^1[h] \circ \varphi) = \Phi_0[\mathrm{S}[h]]$$

and we get $\mathrm{S}^1 = \Phi_0^\dagger \circ \mathrm{D}^1 \circ \Phi_0$ and therefore $\mathrm{S}^\beta = \Phi_0^\dagger \circ \mathrm{D}^\beta \circ \Phi_0$. Exchanging now the differential operator D^β with the weak differential operator $\mathrm{D}^{(\beta)}$, we can show the following intermediate result.

Lemma 2.3.13:
Let $s \in \mathbb{R}_+$, $s \geq 1$, and $\beta \in \mathbb{N}$, $\beta \leq s$, and define $\mathrm{S}_0^{(\beta)} := \Phi_0^\dagger \circ \mathrm{D}_0^{(\beta)} \circ \Phi_0$. Then for any $h \in \mathbb{W}_0^s(\mathbb{R}_+)$,

$$\mathcal{M}_0[\mathrm{S}_0^{(\beta)}[h]](t) = (it)^\beta \mathcal{M}_0[h](t) \text{ for any } t \in \mathbb{R},$$

and $\mathrm{S}_0^{(\beta)} : \mathbb{W}_0^s(\mathbb{R}_+) \to \mathbb{W}_0^{s-\beta}(\mathbb{R}_+)$ is well-defined.

Proof of Lemma 2.3.13. As always, we have $\Phi_0[\mathbb{W}_0^s(\mathbb{R}_+)] = W^s(\mathbb{R}) \subset \mathcal{W}^\beta(\mathbb{R})$ for any $\beta \in \mathbb{N}$, $\beta \leq s$. On this set the operator $\mathrm{D}^{(\beta)}$ is well-defined and we have $\mathcal{F}[\mathrm{D}^{(\beta)}[H]](t) = (it)^\beta \mathcal{F}[H](t)$ for any $t \in \mathbb{R}$ and $H \in \mathcal{W}^\beta(\mathbb{R})$, compare Lemma 2.6.12. We deduce that

$$\mathcal{M}_0[\mathrm{S}_0^{(\beta)}[h]](t) = (2\pi)^{1/2}\mathcal{F}[\mathrm{D}^{(\beta)}[\Phi_0[h]]](t) = (2\pi)^{1/2}(it)^\beta \mathcal{F}[\Phi_0[h]](t) = (it)^\beta \mathcal{M}_0[h](t).$$

The last equation and $h \in \mathbb{W}_0^s(\mathbb{R}_+)$ implies further that $\|(1 + |t|^2)^{(s-\beta)/2}\mathcal{M}_0[\mathrm{S}_0^{(\beta)}[h]]\|_\mathbb{R} < \infty$. □

Now we generalize the last result for any $c \in \mathbb{R}$. To do so, we start again with the strong derivatives. In fact, let us define $\mathrm{D}_c : C_0^\infty(\mathbb{R}) \to C_0^\infty(\mathbb{R})$ through $\mathrm{D}_c := \Phi_c \circ \mathrm{S} \circ \Phi_c^\dagger$ then we have for $H \in C_0^\infty(\mathbb{R})$

$$\mathrm{D}_c[H] = \Phi_c[\mathrm{S}[x^{-c}H(\varphi^\dagger)]] = c\Phi_c[x^{-c}H(\varphi^\dagger)] + \Phi_c[x^{-c}\mathrm{D}[H](\varphi^\dagger)] = cH + \mathrm{D}[H].$$

Due to this construction we have that for any $h \in C_0^\infty(\mathbb{R}_+)$, $(\Phi_c^\dagger \circ \mathrm{D}_c \circ \Phi_c)[h] = (\Phi_0^\dagger \circ \mathrm{D} \circ \Phi_0)[h] = \mathrm{S}[h]$. Now we can show the following Proposition.

Theorem 2.3.14:
Let $s \in \mathbb{R}_+$, $s \geq 1$, and $\beta \in \mathbb{N}$, $\beta \leq s$. Let us define for $c \in \mathbb{R}$ the operator

$$\mathrm{S}_c^{(\beta)} := \Phi_c^\dagger \circ \mathrm{D}_c^{(\beta)} \circ \Phi_c,$$

where $\mathrm{D}_c^{(\beta)} = \mathrm{D}_c^{(1)} \circ \mathrm{D}_c^{(\beta-1)}$ and $\mathrm{D}_c^{(1)}[H] = cH + \mathrm{D}^{(1)}[H]$ for $H \in \mathcal{W}^\beta(\mathbb{R})$. Then for any $h \in \mathbb{W}_c^s(\mathbb{R}_+)$,

$$\mathcal{M}_c[\mathrm{S}_c^{(\beta)}[h]](t) = (c+it)^\beta \mathcal{M}_c[h](t) \text{ for all } t \in \mathbb{R},$$

and $\mathrm{S}_c^{(\beta)} : \mathbb{W}_c^s(\mathbb{R}_+) \to \mathbb{W}_c^{s-\beta}(\mathbb{R}_+)$ is well-defined.

Proof of Theorem 2.3.14. We show the Theorem for the special case $\beta = 1$. The general case follows by induction. The operator $\mathrm{S}_c^{(1)}$ is well-defined, since $\Phi_c[\mathbb{W}_c^s(\mathbb{R}_+)] = W^s(\mathbb{R}) \subset \mathcal{W}^1(\mathbb{R})$ and $D^{(1)}$ is well-defined for any $H \in \mathcal{W}^1(\mathbb{R})$. Next we have

$$\begin{aligned}\mathcal{M}_c[\mathrm{S}_c^{(1)}[h]](t) &= c\mathcal{M}_c[h](t) + (2\pi)^{1/2}\mathcal{F}[\mathrm{D}^{(1)}\Phi_c[h]](t) \\ &= c\mathcal{M}_c[h](t) + it\mathcal{M}_c[h](t) = (c+it)\mathcal{M}_c[h](t)\end{aligned}$$

for any $t \in \mathbb{R}$. As $h \in \mathbb{W}_c^s(\mathbb{R}_+)$ we deduce $\|(1+|t|^2)^{(s-1)/2}\mathcal{M}_c[\mathrm{S}_c^{(1)}[h]]\|_{\mathbb{R}}^2 < \infty$. □

2.4 The Mellin transform for multivariate functions

In this section, our aim is to generalize the definition of Mellin transforms for multivariate functions as it is done by Werner (2006) in the case of Fourier transforms. To do so, we denote for $p \in \mathbb{R}_+, d \in \mathbb{N}, \Omega \subset \mathbb{R}^d$ and for a weight function $\omega : \mathbb{R}^d \to \mathbb{R}_+$ the set $\mathbb{L}^p(\Omega, \omega)$ the quotient set of all p-integrable weighted, complex-valued functions, that is

$$\mathbb{L}^p(\Omega, \omega) := \{h : \Omega \to \mathbb{C} \text{ measurable } : \|h\|^p_{\mathbb{L}^p(\Omega,\omega)} := \int_\Omega |f(\boldsymbol{x})|^p \omega(\boldsymbol{x}) d\boldsymbol{x} < \infty\}/\sim$$

where we define the equivalence relation $\sim$ for two measurable functions $h_1, h_2 : \Omega \to \mathbb{C}$ by $h_1 \sim h_2 :\Leftrightarrow h_1 - h_2 = 0$ Lebesgue-almost everywhere. For the special case $p = 2$ we endow the space $\mathbb{L}^2(\Omega, \omega)$ with its inner product, defined by $\langle h_1, h_2 \rangle_{\mathbb{L}^2(\Omega,\omega)} := \int_\Omega h_1(\boldsymbol{x}) h_2(\boldsymbol{x}) \omega(\boldsymbol{x}) d\boldsymbol{x}$. If $\omega \equiv 1$, we will drop the argument ω and set $\mathbb{L}^p(\Omega) := \mathbb{L}^p(\Omega, 1)$.
For two vectors $\boldsymbol{u} = (u_1, \dots, u_d)^T, \boldsymbol{v} = (v_1, \dots, v_d)^T \in \mathbb{R}^d$ such that there exists no index $i \in [\![d]\!]$ with $v_i = 0$ and $u_i \leq 0$ we define the multivariate potence through

$$\boldsymbol{v}^{\underline{\boldsymbol{u}}} := \prod_{j \in [\![d]\!]} v_j^{u_j}.$$

Additionally, we define the component-wise multiplication $\boldsymbol{uv} := (u_1 v_1, \dots, u_d v_d)$ and if there exists no index $i \in [\![d]\!]$ such that $v_i = 0$ the component-wise division $\frac{\boldsymbol{u}}{\boldsymbol{v}} := \boldsymbol{u}/\boldsymbol{v} := (u_1/v_1, \dots, u_d/v_d)^T$. Further, we denote the usual Euclidean inner product and norm on $\mathbb{R}^d$ through $\langle \boldsymbol{u}, \boldsymbol{v} \rangle_{\mathbb{R}^d} := \sum_{j \in [\![d]\!]} u_i v_i$ and $|\boldsymbol{u}|_{\mathbb{R}^d} := \sqrt{\langle \boldsymbol{u}, \boldsymbol{u} \rangle_{\mathbb{R}^d}}$.

2.4.1 Multiplicative convolution

In contrary to the univariate case, let us consider another convenient approach to define the multiplicative convolution using the additive convolution presented in Definition 2.6.7. Let us define $\varphi : \mathbb{R}^d \to \mathbb{R}^d_+, \boldsymbol{x} \mapsto (\exp(-x_1), \dots, \exp(-x_d))^T$ and denote by $\varphi^\dagger : \mathbb{R}^d_+ \to \mathbb{R}^d, \boldsymbol{x} \mapsto (-\log(x_1), \dots, -\log(x_d))^T$ its inverse. Further, define the isomorphism $\Phi_{\boldsymbol{c}} : \mathbb{L}^p(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{p\boldsymbol{c}-\boldsymbol{1}}}) \to \mathbb{L}^p(\mathbb{R}^d)$ for $\boldsymbol{c} \in \mathbb{R}^d$ and $p \geq 1$ through

$$\Phi_{\boldsymbol{c}}[h] : \boldsymbol{x} \mapsto \varphi(\boldsymbol{x})^{\underline{\boldsymbol{c}}} h(\varphi(\boldsymbol{x}))$$

and denote by $\Phi^\dagger_{\boldsymbol{c}} : \mathbb{L}^p(\mathbb{R}^d) \to \mathbb{L}^p(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{p\boldsymbol{c}-\boldsymbol{1}}})$. The well-definedness of this operator follows by a single change of variable, as seen in the univariate case presented in the last section for $p = 2$. The proof is thus omitted. Now using the operator $\Phi_{\boldsymbol{c}}$ and the additive convolution, we see that the function $(h_1 * h_2) := \Phi^\dagger_{\boldsymbol{c}}[\Phi_{\boldsymbol{c}}[h_1] *_+ \Phi_{\boldsymbol{c}}[h_2]] \in \mathbb{L}^1(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{\boldsymbol{c}-\boldsymbol{1}}})$ for two functions $h_1, h_2 \in \mathbb{L}^1(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{\boldsymbol{c}-\boldsymbol{1}}})$. More precisely, for $\boldsymbol{y} \in \mathbb{R}^d_+$

$$(h_1 * h_2)(\boldsymbol{y}) := \int_{\mathbb{R}^d} h_1(\boldsymbol{y}/\varphi(\boldsymbol{x})) h_2(\varphi(\boldsymbol{x})) d\lambda^d(\boldsymbol{x}) = \int_{\mathbb{R}^d_+} h_1(\boldsymbol{y}/\boldsymbol{x}) h_2(\boldsymbol{x}) \boldsymbol{x}^{\underline{-\boldsymbol{1}}} d\lambda^d(\boldsymbol{x}).$$

Again, by the last equation, we see that the definition of $(h_1 * h_2)$ is independent of the value of $\boldsymbol{c} \in \mathbb{R}^d$.

Definition 2.4.1 (Multiplicative convolution for multivariate functions):
Let $\boldsymbol{c} \in \mathbb{R}^d$ and $h_1, h_2 \in \mathbb{L}^1(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{\boldsymbol{c}-\boldsymbol{1}}})$. Then we call the function $h_1 * h_2 \in \mathbb{L}^1(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{\boldsymbol{c}-\boldsymbol{1}}})$ with

$$(h_1 * h_2)(\boldsymbol{y}) = \int_{\mathbb{R}_+^d} h_1(\boldsymbol{y}/\boldsymbol{x}) h_2(\boldsymbol{x}) \underline{\boldsymbol{x}^{-\boldsymbol{1}}} d\lambda^d(\boldsymbol{x}) \quad \text{for any } \boldsymbol{y} \in \mathbb{R}_+^d,$$

the Multiplicative Convolution of h_1 and h_2.

If we additionally to $h_1, h_2 \in \mathbb{L}^1(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{\boldsymbol{c}-\boldsymbol{1}}})$ assume that $h_1 \in \mathbb{L}^2(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{p\boldsymbol{c}-\boldsymbol{1}}})$ for $p \in [1, \infty]$, we get that $\Phi_{\boldsymbol{c}}[h_1], \Phi_{\boldsymbol{c}}[h_2] \in \mathbb{L}^1(\mathbb{R}^d)$ and $\Phi_{\boldsymbol{c}}[h_1] \in \mathbb{L}^p(\mathbb{R}^d)$. This already implies, that $\Phi_{\boldsymbol{c}}[h_1] *_+ \Phi_{\boldsymbol{c}}[h_2] \in \mathbb{L}^1(\mathbb{R}^d) \cap \mathbb{L}^p(\mathbb{R}^d)$, compare Proposition 2.6.8. More generally spoken, for $h_1 \in \mathbb{L}^1(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{\boldsymbol{c}-\boldsymbol{1}}})$ the convolution operator

$$\bullet * h_1 : \mathbb{L}^p(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{p\boldsymbol{c}-\boldsymbol{1}}}) \cap \mathbb{L}^1(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{\boldsymbol{c}-\boldsymbol{1}}}) \to \mathbb{L}^p(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{p\boldsymbol{c}-\boldsymbol{1}}}) \cap \mathbb{L}^1(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{\boldsymbol{c}-\boldsymbol{1}}}), h_2 \quad \mapsto h_2 * h_1$$

is well-defined with bounded operator norm, which motivates the following multivariate version of Lemma 2.3.1.

Lemma 2.4.2:
Let $h_1, h_2 \in \mathbb{L}^1(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{\boldsymbol{c}-\boldsymbol{1}}})$ and additionally $h_1 \in \mathbb{L}^p(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{p\boldsymbol{c}-\boldsymbol{1}}})$ for $p \in [1, \infty]$. Then $h_1 * h_2 \in \mathbb{L}^1(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{\boldsymbol{c}-\boldsymbol{1}}}) \cap \mathbb{L}^p(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{p\boldsymbol{c}-\boldsymbol{1}}})$ with

$$\|h_1 * h_2\|_{\mathbb{L}^p(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{p\boldsymbol{c}-\boldsymbol{1}}})} \leq \|h_1\|_{\mathbb{L}^p(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{p\boldsymbol{c}-\boldsymbol{1}}})} \|h_2\|_{\mathbb{L}^1(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{\boldsymbol{c}-\boldsymbol{1}}})}.$$

2.4.2 The Mellin transform

We will now define first the Mellin transform for multivariate integrable function, derive another version of the convolution theorem for multivariate function and consider then the Mellin transform for multivariate square-integrable functions.

Definition 2.4.3 (The Mellin transform for integrable functions):
Let $\boldsymbol{c} \in \mathbb{R}^d$ and $h \in \mathbb{L}^1(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{\boldsymbol{c}-\boldsymbol{1}}})$. Then we define the Mellin transform of h developed in $\boldsymbol{c}$ as the function $\mathcal{M}_{\boldsymbol{c}}[h] : \mathbb{R}^d \to \mathbb{C}$ with

$$\mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t}) := \int_{\mathbb{R}_+^d} \underline{\boldsymbol{x}^{\boldsymbol{c}-\boldsymbol{1}+i\boldsymbol{t}}} h(\boldsymbol{x}) d\lambda^d(\boldsymbol{x}) \quad \text{for any } \boldsymbol{t} \in \mathbb{R}^d.$$

Example 2.4.4 (Independent variates):
Let $\boldsymbol{X} = (X_1, X_2)$ be a random vector on $\mathbb{R}_+^2$ where $X_1 \sim f_1$ and $X_2 \sim f_2$ are stochastically independent. Then $\boldsymbol{X}$ possesses the density $f(x_1, x_2) = f_1(x_2) f_2(x_2)$, $x_1, x_2 \in \mathbb{R}_+$, and for $\boldsymbol{c} \in \mathbb{R}^2$ with $\mathbb{E}_f(\underline{\boldsymbol{X}^{\boldsymbol{c}-\boldsymbol{1}}}) = \mathbb{E}_{f_1}(X_1^{c_1-1}) \mathbb{E}_{f_2}(X_2^{c_2-1}) < \infty$ we have

$$\mathcal{M}_{\boldsymbol{c}}[f](\boldsymbol{t}) = \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} x_1^{c_1-1+it_1} x_2^{c_2-1+it_2} f(x_1, x_2) d\lambda(x_1) d\lambda(x_2) = \mathcal{M}_{c_1}[f_1](t_1) \mathcal{M}_{c_2}[f_2](t_2),$$

$\boldsymbol{t} \in \mathbb{R}^2$. More general, if the function $f \in \mathbb{L}^1(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{\boldsymbol{c}-\boldsymbol{1}}})$ can be written as a product of univariate, measurable functions $(f_i)_{i \in [\![d]\!]}$, that is $f(\boldsymbol{x}) = \prod_{i \in [\![d]\!]} f_i(x_i)$, $\boldsymbol{x} \in \mathbb{R}_+^d$, then $f_i \in \mathbb{L}^1(\mathbb{R}_+, x^{c_1-1})$ and

$$\mathcal{M}_{\boldsymbol{c}}[f](\boldsymbol{t}) = \prod_{i \in [\![d]\!]} \mathcal{M}_{c_i}[f_i](t_i), \quad \text{for } \boldsymbol{t} \in \mathbb{R}^d.$$

Example 2.4.5 (Multivariate Log-Normal distribution):
Let us consider the family of multivariate Log-Normal distributions $(\mathbb{P}_{\boldsymbol{\mu},\boldsymbol{\Sigma}})_{(\boldsymbol{\mu},\boldsymbol{\Sigma})\in\mathbb{R}^d\times\mathbb{R}^{(d,d)}_{>0}}$, where $\mathbb{R}^{(d,d)}_{>0}$ denotes the set of symmetric, positive definite $d \times d$-matrices, with densities

$$f_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{x}) = \frac{\boldsymbol{x}^{\underline{-\mathbf{1}}}}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\log(\boldsymbol{x}) - \mu)^T\boldsymbol{\Sigma}^{-1}(\log(\boldsymbol{x}) - \boldsymbol{\mu})\right), \quad \boldsymbol{x} \in \mathbb{R}^d_+,$$

where $\log(\boldsymbol{x}) := (\log(x_1), \dots \log(x_d))^T$. Then for all $\boldsymbol{c} \in \mathbb{R}^d$ we have $f_{\boldsymbol{\mu},\boldsymbol{\Sigma}} \in \mathbb{L}^1(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{\boldsymbol{c}-\mathbf{1}}})$. The Mellin transform of the multivariate Log-Normal distribution is then given by

$$\mathcal{M}_{\boldsymbol{c}}[f_{\boldsymbol{\mu},\boldsymbol{\Sigma}}](\boldsymbol{t}) = \exp(\boldsymbol{\mu}^T(\boldsymbol{c} - \mathbf{1} + i\boldsymbol{t}))\exp\left(\frac{1}{2}(\boldsymbol{c} - \mathbf{1} + i\boldsymbol{t})^T\boldsymbol{\Sigma}(\boldsymbol{c} - \mathbf{1} + i\boldsymbol{t})\right), \quad \boldsymbol{t} \in \mathbb{R}^d.$$

The proof of the following version of the convolution theorem, Proposition 2.2.5, can be done analogously to the univariate version. We therefore omit it.

Proposition 2.4.6 (Convolution theorem for $\mathbb{R}^d_+$- functions):
Let $h_1, h_2 \in \mathbb{L}^1(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{\boldsymbol{c}-\mathbf{1}}})$ for a $\boldsymbol{c} \in \mathbb{R}$. Then,

$$\mathcal{M}_{\boldsymbol{c}}[h_1 * h_2] = \mathcal{M}_{\boldsymbol{c}}[h_1] \cdot \mathcal{M}_{\boldsymbol{c}}[h_2].$$

Let us now define the Mellin transform for arbitrary $h \in \mathbb{L}^2(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{2\boldsymbol{c}-\mathbf{1}}})$ again via the Fourier-Plancherel transform.

Definition 2.4.7 (The Mellin transform for square integrable functions):
Let $\boldsymbol{c} \in \mathbb{R}^d$. For $h \in \mathbb{L}^2(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{2\boldsymbol{c}-\mathbf{1}}})$ we define the Mellin transform developed in $\boldsymbol{c}$ as the function $\mathcal{M}_{\boldsymbol{c}}[h] : \mathbb{R}^d \to \mathbb{C}$ given by

$$\mathcal{M}_{\boldsymbol{c}}[h] := (2\pi)^{d/2}\mathcal{F} \circ \Phi_{\boldsymbol{c}}[h].$$

Similar to the univariate case, for $h \in \mathbb{L}^1(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{\boldsymbol{c}-\mathbf{1}}}) \cap \mathbb{L}^2(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{2\boldsymbol{c}-\mathbf{1}}})$ we have $(2\pi)^{d/2}\mathcal{F}[\Phi_{\boldsymbol{c}}[h]](\boldsymbol{t}) = \int_{\mathbb{R}^d_+} \boldsymbol{x}^{\underline{\boldsymbol{c}-\mathbf{1}+i\boldsymbol{t}}} h(\boldsymbol{x}) d\boldsymbol{x}$ for any $\boldsymbol{t} \in \mathbb{R}^d$. Hence Definition 2.4.7 is an extension of the Mellin transform on $\mathbb{L}^2(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{2\boldsymbol{c}-\mathbf{1}}})$. We derive major properties of the Mellin transform using the theory of Fourier transform as done in the univariate case. Due to these strong similarities of the proofs, we only collect the statements and omit the proofs. The used properties of the Fourier transform can be found in Section 2.6. Let us start with a Plancherel identity for the multivariate Mellin transform.

Proposition 2.4.8 (Plancherel identity for the multivariate Mellin transform):
Let $\boldsymbol{c} \in \mathbb{R}^d$. Then for all $h_1, h_2 \in \mathbb{L}^2(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{2\boldsymbol{c}-\mathbf{1}}})$

$$\langle h_1, h_2 \rangle_{\mathbb{L}^2(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{2\boldsymbol{c}-\mathbf{1}}})} = (2\pi)^{-d}\langle \mathcal{M}_{\boldsymbol{c}}[h_1], \mathcal{M}_{\boldsymbol{c}}[h_2]\rangle_{\mathbb{L}^2(\mathbb{R}^d)}.$$

Further, the Parseval identity holds true, that is $\|h_1\|^2_{\mathbb{L}^2(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{2\boldsymbol{c}-\mathbf{1}}})} = (2\pi)^{-d}\|\mathcal{M}_{\boldsymbol{c}}[h_1]\|^2_{\mathbb{L}^2(\mathbb{R}^d)}$.

Although the existence of an inverse of the multivariate Mellin transform is obviously expressed through $\mathcal{M}^\dagger_{\boldsymbol{c}} := (2\pi)^{-d/2}\Phi^\dagger_{\boldsymbol{c}} \circ \mathcal{F}^\dagger$ we state the following generalization of Proposition 2.3.6.

Proposition 2.4.9 (The inverse Mellin transform for multivariate functions)**:**
For $H \in \mathbb{L}^1(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$ we can express the inverse Mellin transform $\mathcal{M}_{\boldsymbol{c}}^{\dagger} : \mathbb{L}^2(\mathbb{R}^d) \to \mathbb{L}^2(\mathbb{R}_+^d, \boldsymbol{x}^{\underline{2\boldsymbol{c}-\boldsymbol{1}}})$ for $\boldsymbol{c} \in \mathbb{R}^d$ of H explicitly by

$$\mathcal{M}_{\boldsymbol{c}}^{\dagger}[H](\boldsymbol{x}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} H(\boldsymbol{t}) \boldsymbol{x}^{\underline{-\boldsymbol{c}-i\boldsymbol{t}}} d\boldsymbol{t}, \qquad \text{for any } \boldsymbol{x} \in \mathbb{R}_+^d.$$

2.4.3 Mellin-Sobolev space

Let us now develop the Sobolev spaces theory for the Mellin transform of multivariate functions. We start with defining the Mellin-Sobolev spaces, derive a Lemma of Sobolev-type result for these spaces and investigate the differential operator on these spaces. In contrary to the univariate case, we additionally define the so-called anisotropic Mellin-Sobolev spaces as another natural generalization of the univariate Mellin-Sobolev spaces.

Definition 2.4.10 (The Mellin-Sobolev space)**:**
For $s \in \mathbb{R}_+$ and $\boldsymbol{c} \in \mathbb{R}^d$ we define the MELLIN-SOBOLEV SPACES through

$$\mathbb{W}_{\boldsymbol{c}}^s(\mathbb{R}_+^d) := \{h \in \mathbb{L}^2(\mathbb{R}_+^d, \boldsymbol{x}^{\underline{2\boldsymbol{c}-\boldsymbol{1}}}) : \|(1 + |\boldsymbol{t}|_{\mathbb{R}^d}^2)^{s/2} \mathcal{M}_{\boldsymbol{c}}[h]\|_{\mathbb{L}^2(\mathbb{R}^d)} < \infty\}.$$

Since $\Phi_{\boldsymbol{c}}[\mathbb{W}_{\boldsymbol{c}}^s(\mathbb{R}_+^d)] = W^s(\mathbb{R}^d)$, we obtain by similar arguments a multivariate version of Theorem 2.3.10.

Theorem 2.4.11 (Lemma of Sobolev for the multivariate Mellin transform)**:**
Let $m, k \in \mathbb{N}_0$ such that $m > k + d/2$. For each equivalence class $[h] \in \mathbb{W}_{\boldsymbol{c}}^m(\mathbb{R}_+^d)$ there exists a k-times continuously differentiable representant of $[h]$.

Let us denote by $C_0^\infty(\mathbb{R}_+^d)$ the set of all smooth functions with compact support in $\mathbb{R}_+^d$. For $j \in [\![d]\!]$ let us define the operator $\mathrm{S}_j : C_0^\infty(\mathbb{R}_+^d) \to C_0^\infty(\mathbb{R}_+^d), h \mapsto -x_j \frac{\partial}{\partial x_j} h$. Further for any vector $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^T \in \mathbb{N}_0^d$ let us define the operator $\mathrm{S}^{\boldsymbol{\beta}}$ by $\mathrm{S}^{\boldsymbol{\beta}} = \mathrm{S}_1^{\beta_1} \circ \cdots \circ \mathrm{S}_d^{\beta_d}$ where $\mathrm{S}_j^{\beta_j} = \mathrm{S}_j \circ \cdots \circ \mathrm{S}_j$ the β_j-times self composition of S_j where $S_j^0 := \mathrm{Id}_{C_0^\infty(\mathbb{R}_+^d)}$ is the identity on $C_0^\infty(\mathbb{R}_+^d)$. Then we can prove the following Proposition.

Proposition 2.4.12:
For any multi index $\boldsymbol{\beta} \in \mathbb{N}_0^d$, $\boldsymbol{c} \in \mathbb{R}^d$ and $h \in C_0^\infty(\mathbb{R}_+^d)$

$$\mathcal{M}_{\boldsymbol{c}}[\mathrm{S}^{\boldsymbol{\beta}} \circ h](\boldsymbol{t}) = (\boldsymbol{c} + i\boldsymbol{t})^{\boldsymbol{\beta}} \mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t}) \quad \text{for any } \boldsymbol{t} \in \mathbb{R}_+^d.$$

Proof of Proposition 2.4.12. It is sufficient to show that for any $j \in [\![d]\!]$ and $\boldsymbol{t} \in \mathbb{R}^d$ we have $\mathcal{M}_{\boldsymbol{c}}[\mathrm{S}_j[h]](\boldsymbol{t}) = (c_j + it_j)\mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t})$. In fact,

$$\mathcal{M}_{\boldsymbol{c}}[\mathrm{S}_j[h]](\boldsymbol{t}) = \int_{\mathbb{R}_+^{d-1}} \boldsymbol{x}_{-j}^{\underline{\boldsymbol{c}_{-j}-\boldsymbol{1}+i\boldsymbol{t}_{-j}}} \int_{\mathbb{R}_+} \left(-x_j^{c_j+it_j} \frac{\partial}{\partial x_j} h(\boldsymbol{x}) \right) d\lambda(x_j) d\lambda^{d-1}(\boldsymbol{x}_{-j})$$

where we denote by $\boldsymbol{v}_{-j}$ for a d-dimensional vector $\boldsymbol{v} \in \mathbb{R}^d$ the $d-1$-dimensional vector $\boldsymbol{v}_{-j} = (v_1, \dots, v_{j-1}, v_{j+1}, \dots v_d)^T$. Analogously to the proof of Proposition 2.3.12, using the integration by parts for the inner integral proves the claim. □

To deduce a weak version of S^{β} for the multivariate case, we define for any $j \in [\![d]\!]$ the operator $\mathrm{D}_{\boldsymbol{c}}^{(\boldsymbol{e}_j)}[H] := c_j H + \mathrm{D}^{(\boldsymbol{e}_j)} H$ for $H \in \mathcal{W}^s(\mathbb{R}^d)$ with $s \geq 1$. Here $\boldsymbol{e_j} \in \mathbb{N}_0^d$ denotes the $j-th$ canonical vector, that is $\boldsymbol{e_j} = (e_{j1}, \dots, e_{jd})^T$ with $e_{ji} = \mathbb{1}_{j=i}$. Then we define for $\boldsymbol{\beta} \in \mathbb{N}_0^d$ and any $s \in \mathbb{R}_+$ with $|\boldsymbol{\beta}|_1 := \sum_{j \in [\![d]\!]} \beta_j \leq s$ the operator $\mathrm{D}_{\boldsymbol{c}}^{(\boldsymbol{\beta})} : \mathcal{W}^s(\mathbb{R}^d) \to \mathcal{W}^{s-|\boldsymbol{\beta}|_1}(\mathbb{R}^d)$ through

$$\mathrm{D}_{\boldsymbol{c}}^{(\boldsymbol{\beta})} = \underbrace{\mathrm{D}_{\boldsymbol{c}}^{(\boldsymbol{e}_1)} \circ \dots \circ \mathrm{D}_{\boldsymbol{c}}^{(\boldsymbol{e}_1)}}_{\beta_1-\text{times}} \circ \dots \circ \underbrace{\mathrm{D}_{\boldsymbol{c}}^{(\boldsymbol{e}_d)} \circ \dots \circ \mathrm{D}_{\boldsymbol{c}}^{(\boldsymbol{e}_d)}}_{\beta_d-\text{times}}$$

and define $\mathrm{S}_{\boldsymbol{c}}^{(\boldsymbol{\beta})} : \mathbb{W}_{\boldsymbol{c}}^{s}(\mathbb{R}_+^d) \to \mathbb{W}_{\boldsymbol{c}}^{s-|\boldsymbol{\beta}|_1}(\mathbb{R}_+^d)$ through $\mathrm{S}_{\boldsymbol{c}}^{\boldsymbol{\beta}} := \Phi_c^{\dagger} \circ \mathrm{D}_{\boldsymbol{c}}^{(\boldsymbol{\beta})} \circ \Phi_{\boldsymbol{c}}$.

Theorem 2.4.13:
Let $s \in \mathbb{R}_+$ and $\boldsymbol{c} \in \mathbb{R}^d$. Then for any $h \in \mathbb{W}_{\boldsymbol{c}}^{s}(\mathbb{R}_+^d)$ and any $\boldsymbol{\beta} \in \mathbb{N}_0^d, |\boldsymbol{\beta}|_1 \leq s$,

$$\mathcal{M}_{\boldsymbol{c}}[\mathrm{S}_{\boldsymbol{c}}^{(\boldsymbol{\beta})} \circ h](\boldsymbol{t}) = (\boldsymbol{c} + i\boldsymbol{t})^{\boldsymbol{\beta}} \mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t}) \quad \text{for any ăt} \in \mathbb{R}^d.$$

Proof of Theorem 2.4.13. It is sufficient to show the theorem for the special case $\boldsymbol{\beta} = \boldsymbol{e}_j$ for $j \in [\![d]\!]$. In fact, we have for $\boldsymbol{t} \in \mathbb{R}^d$,

$$\begin{aligned}
\mathcal{M}_{\boldsymbol{c}}[\mathrm{S}_{\boldsymbol{c}}^{(\boldsymbol{e}_j)}[h]](\boldsymbol{t}) &= c_j \mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t}) + (2\pi)^{d/2} \mathcal{F}[\mathrm{D}^{(\boldsymbol{e}_j)}[\Phi_c[h]]](\boldsymbol{t}) \\
&= c_j \mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t}) + i t_j (2\pi)^{d/2} \mathcal{F}[\Phi_c[h]](\boldsymbol{t}) \\
&= (c_j + i t_j) \mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t}).
\end{aligned}$$

□

Let us now define the anisotropic counterpart to the Mellin-Sobolev spaces defined in Definition 2.4.10.

Definition 2.4.14 (The anisotropic Mellin-Sobolev spaces)**:**
Let $\boldsymbol{s} \in \mathbb{R}_+^d$, $\boldsymbol{c} \in \mathbb{R}^d$ we define the ANISOTROPIC MELLIN-SOBOLEV SPACE through

$$\mathbb{W}_{\boldsymbol{c}}^{\boldsymbol{s}}(\mathbb{R}_+^d) := \{h \in \mathbb{L}^2(\mathbb{R}_+^d, \boldsymbol{x}^{\underline{2\boldsymbol{c}-\boldsymbol{1}}}) : \sum_{j \in [\![d]\!]} \|(1 + |t_j|^2)^{s_j/2} \mathcal{M}_{\boldsymbol{c}}[h]\|_{\mathbb{L}^2(\mathbb{R}^d)}^2 < \infty\}.$$

For the case $\boldsymbol{s} \in \mathbb{N}_0^d$ we can characterize these spaces by the following proposition which is a direct consequence of its Fourier counterpart, Proposition 2.6.14, and its proof is thus omitted. Here $|\boldsymbol{v}|_1 := \sum_{j \in [\![d]\!]} |v_j|$ for $\boldsymbol{v} \in \mathbb{R}^d$.

Proposition 2.4.15:
For $\boldsymbol{s} \in \mathbb{N}_0^d$,

$$\mathbb{W}_{\boldsymbol{c}}^{\boldsymbol{s}}(\mathbb{R}_+^d) = \{h \in \mathbb{L}^2(\mathbb{R}_+^d, \boldsymbol{x}^{\underline{2\boldsymbol{c}-\boldsymbol{1}}}) : \sum_{\substack{\boldsymbol{\gamma} \in \mathbb{N}_0^d \\ |\boldsymbol{\gamma}/\boldsymbol{s}|_1 \leq 1}} \|\mathrm{S}_{\boldsymbol{c}}^{(\boldsymbol{\gamma})}[h]\|_{\mathbb{L}^2(\mathbb{R}_+^d, \boldsymbol{x}^{\underline{2\boldsymbol{c}-\boldsymbol{1}}})} < \infty\}.$$

We can prove the following version of Theorem 2.4.13 for the anisotropic Mellin-Sobolev space.

Theorem 2.4.16:
Let $\boldsymbol{s} \in \mathbb{N}_0^d$, $\boldsymbol{c} \in \mathbb{R}^d$ and $h \in \mathbb{W}_{\boldsymbol{c}}^{\boldsymbol{s}}(\mathbb{R}_+^d)$. Then for every $j \in [\![d]\!]$ we have $\mathrm{S}_{\boldsymbol{c}}^{(s_j \boldsymbol{e_j})}[h] \in \mathbb{L}^2(\mathbb{R}_+^d, \boldsymbol{x}^{\underline{\boldsymbol{2c-1}}})$ and

$$\mathcal{M}_{\boldsymbol{c}}[\mathrm{S}_{\boldsymbol{c}}^{(s_j \boldsymbol{e_j})} \circ h](\boldsymbol{t}) = (c_j + it_j)^{s_j} \mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t}) \quad \text{for any } \boldsymbol{t} \in \mathbb{R}^d.$$

Proof of Theorem 2.4.16. From $|(s_j \boldsymbol{e_j})/\boldsymbol{s}_j|_1 = s_j/s_j = 1$ we can deduce using Proposition 2.4.15 that $\mathrm{S}_{\boldsymbol{c}}^{(s_j \boldsymbol{e_j})}[h] \in \mathbb{L}^2(\mathbb{R}_+^d, \boldsymbol{x}^{\boldsymbol{2c-1}})$. The rest of the proof follows the same steps as Theorem 2.4.13. □

2.5 Approximations and numerical study

In this section we want to use the Mellin transform in Definition 2.4.7 to build approximations of a function $h \in \mathbb{L}^2(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{2c-1}})$ in the $\mathbb{L}^2(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{2c-1}})$-sense. First we will use the inverse of the Mellin transform, Proposition 2.3.6, to construct the approximations. Then we will analyze the upcoming approximation error for functions lying in a Mellin-Sobolev space, respectively an anisotropic Mellin-Sobolev space. We end this section with an illustration by plotting both the true functions $h \in \mathbb{L}^2(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{2c-1}})$ and their approximations for the case $d = 1$ and $d = 2$ for some examples of functions.

2.5.1 Theoretical study

For the upcoming construction, let us assume that $h \in \mathbb{L}^2(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{2c-1}})$ and $\boldsymbol{k} \in \mathbb{R}^d_+$. Let us first define the hypercuboid $[-\boldsymbol{k}, \boldsymbol{k}] := \times_{j \in [\![d]\!]}[-k_j, k_j]$. Since $\mathbb{1}_{[-\boldsymbol{k},\boldsymbol{k}]}\mathcal{M}_{\boldsymbol{c}}[h] \in \mathbb{L}^1(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$ we can define the function $h_{\boldsymbol{k}} \in \mathbb{L}^2(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{2c-1}})$ through $h_{\boldsymbol{k}} := \mathcal{M}^{\dagger}_{\boldsymbol{c}}[\mathcal{M}_{\boldsymbol{c}}[h]\mathbb{1}_{[-\boldsymbol{k},\boldsymbol{k}]}]$. More precisely, applying Proposition 2.3.6, we get for any $\boldsymbol{x} \in \mathbb{R}^d_+$,

$$h_{\boldsymbol{k}}(\boldsymbol{x}) = \frac{1}{(2\pi)^d} \int_{[-\boldsymbol{k},\boldsymbol{k}]} \mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t})\boldsymbol{x}^{-\boldsymbol{c}-i\boldsymbol{t}} d\lambda^d(\boldsymbol{t}). \tag{2.4}$$

It is worth stressing out, that the functions $(h_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathbb{R}^d_+}$ are also dependent on the model parameter $\boldsymbol{c} \in \mathbb{R}^d$. A more precise notation would therefore include the parameter $\boldsymbol{c} \in \mathbb{R}^d$. For the sake of readability we exclude this index and make sure that the value of the model parameter $\boldsymbol{c} \in \mathbb{R}$ is always stated.
We will now consider the upcoming approximation error. In fact, we consider the induced metric on the Hilbert space $\mathbb{L}^2(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{2c-1}})$. The approximation error is then given for any $\boldsymbol{k} \in \mathbb{R}^d_+$ by

$$\|h_{\boldsymbol{k}} - h\|^2_{\mathbb{L}^2(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{2c-1}})} = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d \setminus [-\boldsymbol{k},\boldsymbol{k}]} |\mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t})$$

applying the Parseval identity, Proposition 2.4.8. Obviously, for $\boldsymbol{k} \to \infty$, where the limit is understood component-wise, the approximation error goes to 0.
More precisely, if we assume that h lies in the Mellin-Sobolev space $\mathbb{W}^s_{\boldsymbol{c}}(\mathbb{R}^d)$ for $s \in \mathbb{R}_+$, see Definition 2.4.10, we get

$$\begin{aligned}\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d \setminus [-\boldsymbol{k},\boldsymbol{k}]} |\mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d \setminus [-\boldsymbol{k},\boldsymbol{k}]} (1 + |\boldsymbol{t}|^2)^s (1 + |\boldsymbol{t}|^2)^{-s} |\mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}) \\ &\leq (\min(k_1, \dots, k_d))^{-2s} \, \|(1 + |\boldsymbol{t}|^2)^{s/2}\mathcal{M}_{\boldsymbol{c}}[h]\|^2_{\mathbb{L}^2(\mathbb{R}^d)}\end{aligned}$$

If we assume that h lies in an anisotropic Mellin-Sobolev space, that is $h \in \mathbb{W}^{\boldsymbol{s}}_{\boldsymbol{c}}(\mathbb{R}^d)$ for a $\boldsymbol{s} \in \mathbb{R}^d_+$, see Definition 2.4.14, we get by using $\bigcup_{j\in[\![d]\!]}\{\boldsymbol{t} \in \mathbb{R}^d : |t_j| > k_j\} \supseteq \mathbb{R}^d \setminus [-\boldsymbol{k}, \boldsymbol{k}]$ that

$$\begin{aligned}\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d\setminus[-\boldsymbol{k},\boldsymbol{k}]} |\mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}) &\leq \sum_{j=1}^d \int_{\mathbb{R}^d} \mathbb{1}_{|t_j|>k_j} |\mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}) \\ &\leq \sum_{j=1}^d \int_{\mathbb{R}^d} \mathbb{1}_{|t_j|>k_j}(1+|t_j|^2)^{s_j}(1+|t_j|^2)^{-s_j} |\mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t})|^2 d\lambda^d(\boldsymbol{t}) \\ &\leq \left(\sum_{j=1}^d k_j^{-2s_j}\right)\left(\sum_{j=1}^d \|(1+|t_j|^2)^{s_j/2}\mathcal{M}_{\boldsymbol{c}}[h]\|^2_{\mathbb{L}^2(\mathbb{R}^d)}\right).\end{aligned}$$

For the case of $d = 1$, these two spaces are equal. As we can see, a higher assumption on the regularity is directly connected to the decay of the Mellin transform and thus on the decay of the approximation error.

Univariate case Let us consider the following four examples

- (i) Beta-Distribution: $h(x) := B(2,a)^{-1}x(1-x)^{a-1}\mathbb{1}_{(0,1)}(x)$ for $a \in \mathbb{N}$,
- (ii) Log-Gamma Distribution: $h(x) = \frac{1}{\Gamma(b)}x^{-2}\log(x)^{b-1}\mathbb{1}_{(1,\infty)}(x)$ for $b \in \mathbb{R}_+$;
- (iii) Gamma-Distribution: $h(x) = \frac{1}{\Gamma(e)}x^{e-1}\exp(-x)\mathbb{1}_{(0,\infty)}(x)$ for $e \in \mathbb{R}_+$ and
- (iv) Log-Normal Distribution: $h(x) = \frac{1}{\sqrt{2f\pi x}}\exp(-\log(x)^2/(2f))\mathbb{1}_{\mathbb{R}_+}(x)$ for $f \in \mathbb{R}_+$.

We will consider the three different cases of $c = 0$, $c = 1/2$ and $c = 1$. In the first case, the weight function is given by $x \mapsto x^{-1}$, for the second case by $x \mapsto 1$ and for the third through $x \mapsto x$. To ensure that all examples can be approximated for all three values of c, we need to assume additionally $e > 1$ for the Gamma distribution. The Mellin transform of the four functions are then given in

- (i) Example 2.1.6 by $\mathcal{M}_c[h](t) = \prod_{j=1}^a \frac{j+1}{c+j+it}$, $t \in \mathbb{R}$,
- (ii) Example 2.1.10 by $\mathcal{M}_c[h](t) = (2-c-it)^{-b}$, $t \in \mathbb{R}$,
- (iii) Example 2.1.7 by $\mathcal{M}_c[h](t) = \Gamma(e+c-1+it)/\Gamma(d)$, $t \in \mathbb{R}$, and
- (iv) Example 2.1.11 by $\mathcal{M}_c[h](t) = \exp(f(c-1+it)^2/2)$, $t \in \mathbb{R}$.

Let us now consider when the examples lay in the Mellin-Sobolev space. For the cases (i), (ii) and (iv) this can be directly derived from the Mellin transform. For the case (iii) we needs to apply the Stirling formula, see Andrews et al. (1999), which states that $|\Gamma(\sigma+it)| \leq |t|^{\sigma-1/2}\exp(-\pi|t|/2)$ for $|t| \geq 2$ and $\sigma \geq -2$. We get

- (i) $h \in \mathbb{W}^s_c(\mathbb{R}_+)$ if and only if $s < a - 1/2$,
- (ii) $h \in \mathbb{W}^s_c(\mathbb{R}_+)$ if and only if $s < b - 1/2$,
- (iii) $h \in \mathbb{W}^s_c(\mathbb{R}_+)$ for all $s \in \mathbb{R}_+$ and
- (iv) $h \in \mathbb{W}^s_c(\mathbb{R}_+)$ for all $s \in \mathbb{R}_+$.

More precisely, the approximation error is decaying with an exponential rate in the cases (iii) and (iv). We will plot the approximations and the true densities to visualize the approximation error and the effects of the choice of $c \in \{0, 0.5, 1\}$.

Bivariate case Let us consider the examples

(a) Beta/Log-Normal distribution: $h(\boldsymbol{x}) = h_1(x_1)h_2(x_2)$ with $h_1(x_1) = B(2,a)^{-1}x_1(1-x_1)^{a-1}\mathbb{1}_{(0,1)}(x_1)$ and $h_2(x_2) = \frac{1}{\sqrt{2\pi b}x_2}\exp(-\frac{\log(x_2)^2}{2b})\mathbb{1}_{\mathbb{R}_+}(x_2)$ for $a \in \mathbb{N}$ and $b > 0$;

(b) Bivariate-Log-Normal distribution: $h(\boldsymbol{x}) = \frac{1}{2\pi x_1 x_2 |\boldsymbol{\Sigma}|^{1/2}}\exp(-\frac{1}{2}(\log(\boldsymbol{x}))^T\boldsymbol{\Sigma}^{-1}\log(\boldsymbol{x}))$ for $\boldsymbol{\Sigma}$ symmetric positiv-definit matrix in $\mathbb{R}^{2\times 2}$.

Then the Mellin transforms for $\boldsymbol{c} \in \{0.1/2, 1\}^2$ are given by

(a) $\mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t}) = \exp(\frac{b}{2}(c_2 - 1 + it_2)^2)\prod_{j=1}^{a}\frac{j+1}{c_1+j+it_1}$, for $\boldsymbol{t} \in \mathbb{R}^2$ and

(b) $\mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t}) = \exp(\frac{1}{2}(\boldsymbol{c} - \mathbf{1} + i\boldsymbol{t})^T\boldsymbol{\Sigma}(\boldsymbol{c} - \mathbf{1} + i\boldsymbol{t}))$ for $\boldsymbol{t} \in \mathbb{R}^2$.

Now we study whether the examples are in the isotropic Mellin Sobolev-space $\mathbb{W}^s_{\boldsymbol{c}}(\mathbb{R}^2_+)$, respectively the anisotropic Mellin-Sobolev space $\mathbb{W}^{\boldsymbol{s}}_{\boldsymbol{c}}(\mathbb{R}^2_+)$. Indeed, we have

(a) $h \in \mathbb{W}^s_{\boldsymbol{c}}(\mathbb{R}^2_+)$ if and only if $s < a - 1/2$, while $h \in \mathbb{W}^{\boldsymbol{s}}_{\boldsymbol{c}}(\mathbb{R}^2_+)$ if and only if for $\boldsymbol{s} \in (0, a - 1/2) \times \mathbb{R}_+$ arbitrary;

(b) $h \in \mathbb{W}^s_{\boldsymbol{c}}(\mathbb{R}^2_+)$ for all $s \in \mathbb{R}_+$, respectively $h \in \mathbb{W}^{\boldsymbol{s}}_{\boldsymbol{c}}(\mathbb{R}^2_+)$, $\boldsymbol{s} \in \mathbb{R}^2_+$.

It can be seen that the anisotropic Mellin-Sobolev space captures more detailed the decay properties of the Mellin transform in comparison to the Mellin-Sobolev space. Indeed, in example (a) it can be nicely seen that the regularity parameter $s \in \mathbb{R}_+$ corresponds to the slowest decay in one of the directions of the Mellin transform, while the regularity parameter $\boldsymbol{s} \in \mathbb{R}^2_+$ can better distinguish in this case the possibly different decay in different directions.

2.5.2 Numerical study

We start by comparing all examples with varying parameter $k \in \mathbb{R}_+$ for the case $d = 1$ and $c = 1/2$. Comparing Figure 2.1 and Figure 2.2 we can see that the approximation error seems to decay much faster for the examples (iii) and (iv) as for the examples (i) and (ii). This is a behavior which is predicted by the theoretical results.

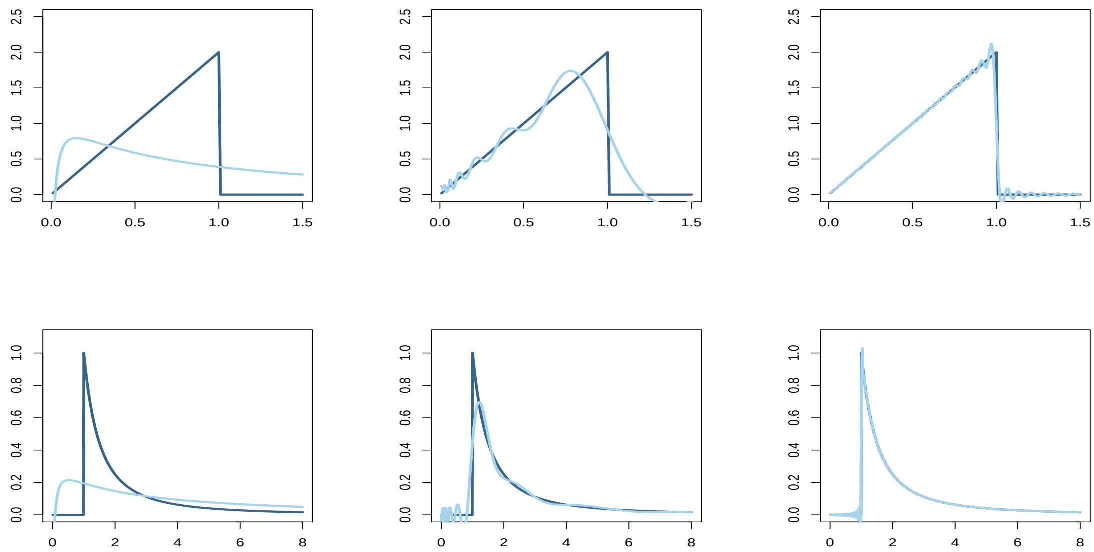

Figure 2.1: Considering the examples (i) (top) and (ii) (bottom) with $c = 1/2$ with $k = 1$ (left), $k = 10$ (middle) and $k = 100$ (right). For the case (i) we set $a = 1$ and for (ii) we set $b = 1$. The dark blue curve is the true function h while the function h_k is represented by the bright blue curve.

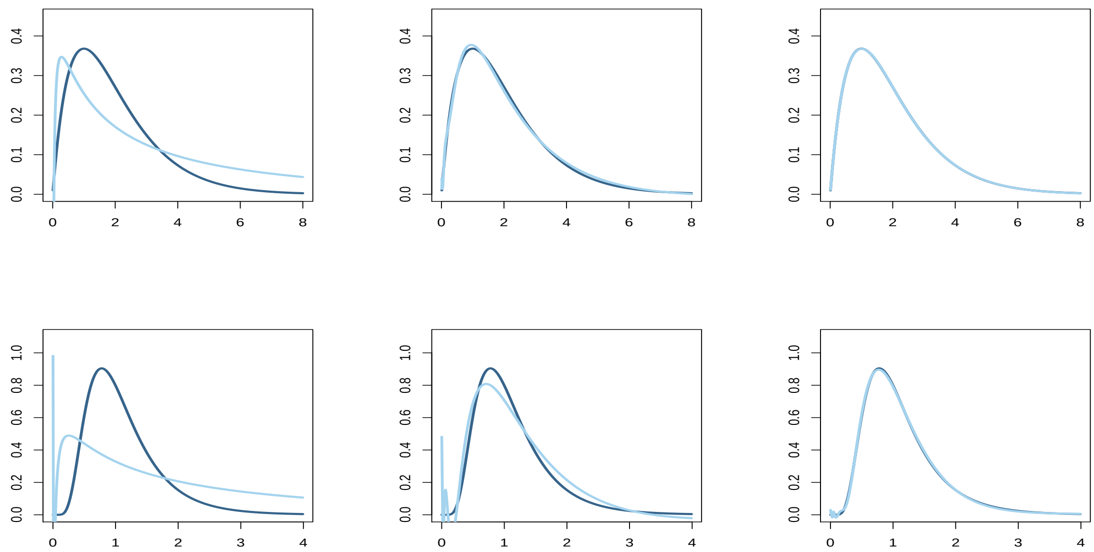

Figure 2.2: Considering the examples (iii) (top) and (iv) (bottom) with $c = 1/2$ with $k = 1$ (left), $k = 3$ (middle) and $k = 5$ (right). For the case (iii) we set $e = 2$ and for (iv) we set $f = 0.25$. The dark blue curve is the true function h while the function h_k is represented by the bright blue curve.

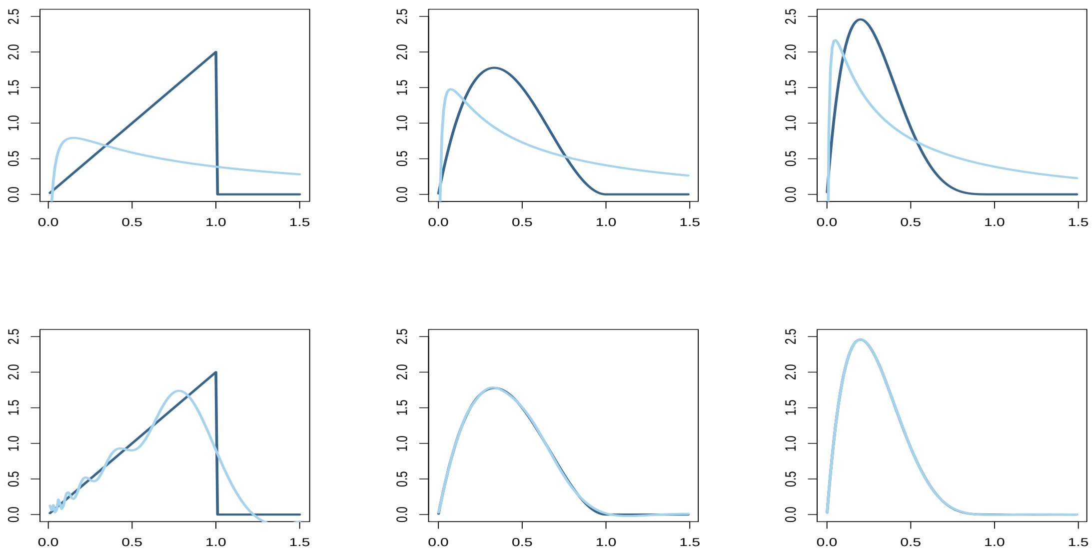

Figure 2.3: Considering the examples (i) with $b = 1$ (left), $b = 3$ (middle) and $b = 5$ (right) with $c = 1/2$. Here we displayed the approximations with $k = 1$ (top) and $k = 10$ (bottom). The dark blue curve is the true function h while the function h_k is represented by the bright blue curve.

Now let us consider example (i) for increasing parameters $b \in \mathbb{N}$, keeping in mind that the parameter determines in which Mellin-Sobolev space the functions lays in. As the theory would suggest for higher values of $b \in \mathbb{N}$ the function h lies in Mellin-Sobolev space with higher smoothness parameter $s \in \mathbb{R}_+$ which on the other hand implies that the decay of the approximation error is stronger, Figure 2.3 seems to imply that higher values of $b \in \mathbb{N}$ implies a better approximation through the Mellin transform.

To close the simulation for the univariate case, we will now consider the case of (iv) with varying value of $c \in \{0, 0.5, 1\}$. Based on Figure 2.4 it seems that the approximations differ in their behavior with respect to the location on the positive real line. Meaning that for values of $c \in \mathbb{R}$ smaller than $1/2$, the point-wise distance of the approximation seems to be smaller for x-values close to 0 while for values of $c \in \mathbb{R}$ bigger than $1/2$ the opposite behavior seem to occur. For $c = 1/2$ neither of those effects seems to occur. The bottom row of Figure 2.4 suggests that the mentioned effects seems to be negligible for higher values of $k \in \mathbb{R}_+$.

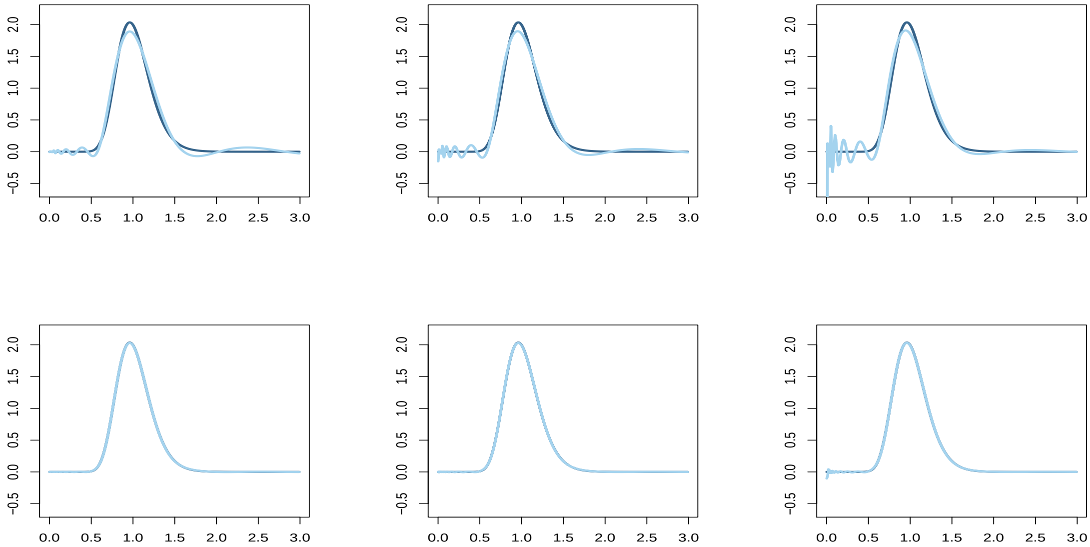

Figure 2.4: Considering the example (iv) with $c = 0$ (left), $c = 0.5$ (middle) and $c = 1$ (right) with $f = 0.04$. Here we displayed the approximations with $k = 9$ (top) and $k = 15$ (bottom). The dark blue curve is the true function h while the function h_k is presented by the bright blue curve.

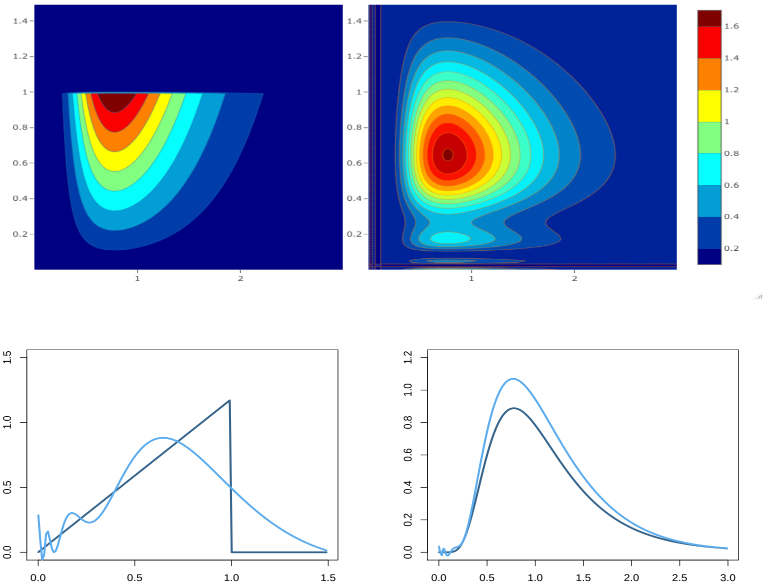

Figure 2.5: Considering the example (a) with $a = 1$, $b = 0.2$ with the true density (left-top) compared with the approximation $h_{\boldsymbol{k}}$ (right, top) for $\boldsymbol{k} = (5,5)^T$ and $\boldsymbol{c} = (1/2, 1/2)^T$. Here we display the intersection of the graph in the bottom row, where the bright blue curve is the approximation and the dark blue curve is the true density.

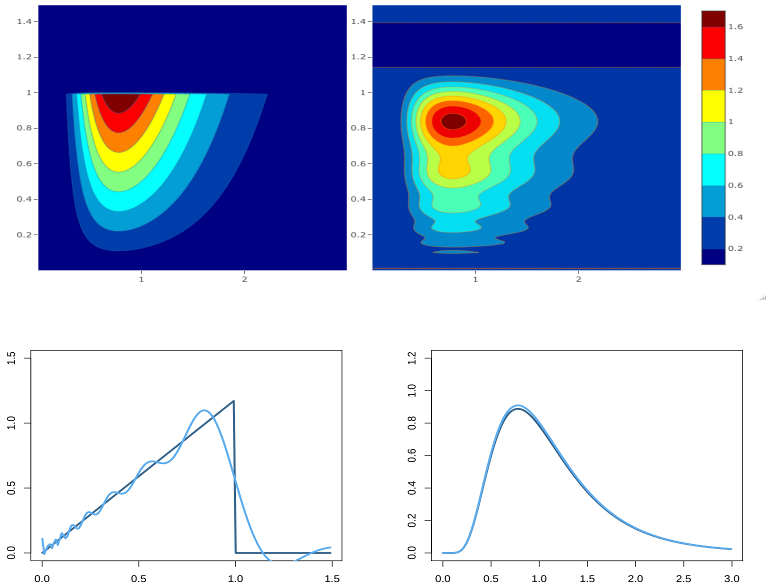

Figure 2.6: Considering the example (a) with $a = 1$, $b = 0.2$ with the true density (left-top) compared with the approximation $h_{\boldsymbol{k}}$ (right, top) for $\boldsymbol{k} = (15, 15)^T$ and $\boldsymbol{c} = (1/2, 1/2)^T$. Here we display the intersection of the graph in the bottom row, where the bright blue curve is the approximation and the dark blue curve is the true density.

Comparing the plots Figure 2.5 and Figure 2.6, we see that the approximation in the direction of the Log-Normal distribution works much better than the approximation of the Beta distribution direction. Indeed, we will see that as soon as it comes to estimation of the density, we will have an additional error term due to the estimation step, which will be increasing in each direction k_1, k_2 of $\boldsymbol{k} \in \mathbb{R}_+^2$ making it necessary to carefully choose the values of k_1, k_2 and mostly consider anisotropic choices, that is $k_1 \neq k_2$, motivated by the plots Figure 2.5, respectively Figure 2.6.

2.6 Appendix: The Fourier transform and the Sobolev space

In this section, we collect the used results of the theory of Fourier transform and Sobolev spaces. For a detailed introduction to this theory see Werner (2006) and Triebel (2006).

2.6.1 The Fourier transform

Definition 2.6.1 (Fourier transform of $\mathbb{L}^1(\mathbb{R}^d)$)**:**
For $H \in \mathbb{L}^1(\mathbb{R}^d)$ we define the Fourier transform $\mathcal{F}[H] : \mathbb{R}^d \to \mathbb{C}$ of H by

$$\mathcal{F}[H](\boldsymbol{t}) := (2\pi)^{-d/2} \int_{\mathbb{R}^d} H(\boldsymbol{x}) \exp(-i\langle \boldsymbol{x}, \boldsymbol{t}\rangle_{\mathbb{R}^d}) d\lambda^d(\boldsymbol{x})$$

for any $\boldsymbol{t} \in \mathbb{R}^d$.

Let us now define the Fourier transform of a square integrable function, that is $H \in \mathbb{L}^2(\mathbb{R}^d)$. To do so, we will first introduce the so-called Schwartz space which lies dense in all $\mathbb{L}^p(\mathbb{R}^d)$ spaces for $1 \leq p < \infty$. On this space we will consider Fourier transform for $\mathbb{L}^1$-functions, show that the operator maps to $\mathbb{L}^2(\mathbb{R}^d)$ and is bounded. From this, we derive an extension on the whole set $\mathbb{L}^2(\mathbb{R}^d)$ which will be referred to as the Fourier-Plancherel transform.
The Schwartz spaces is defined by

$$\mathscr{S}(\mathbb{R}^d) := \{f \in C^\infty(\mathbb{R}^d) : \forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}_0^d : \lim_{|\boldsymbol{x}| \to 0} \boldsymbol{x}^{\boldsymbol{\alpha}} \mathrm{D}^{\boldsymbol{\beta}} f(\boldsymbol{x}) = 0\}$$

where $C^\infty(\mathbb{R}^d)$ is the set of all smooth function on $\mathbb{R}^d$. The following Theorem is a classical result of the Fourier analysis. A proof can be found in Werner (2006).

Theorem 2.6.2:
The Fourier transform $\mathcal{F} : \mathscr{S}(\mathbb{R}^d) \to \mathscr{S}(\mathbb{R}^d)$ is an automorphism with inverse $\mathcal{F}^\dagger : \mathscr{S}(\mathbb{R}^d) \to \mathscr{S}(\mathbb{R}^d)$ given

$$\mathcal{F}^\dagger[f](\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^n} f(\boldsymbol{t}) \exp(i\langle \boldsymbol{x}, \boldsymbol{t}\rangle_{\mathbb{R}^d}) d\lambda^d(\boldsymbol{t}), \qquad \text{for any} \boldsymbol{x} \in \mathbb{R}^n, f \in \mathscr{S}(\mathbb{R}^d).$$

Further, the Plancherel identity holds true, that is for all $f, h \in \mathscr{S}(\mathbb{R}^d)$, $\langle \mathcal{F}[f], \mathcal{F}[g]\rangle_{\mathbb{L}^2(\mathbb{R}^d)} = \langle f, g\rangle_{\mathbb{L}^2(\mathbb{R}^d)}$.

From the last theorem, we deduce that $\mathcal{F}$ is well-defined on the dense subspace $\mathscr{S}(\mathbb{R}^d) \subset \mathbb{L}^2(\mathbb{R}^d)$, isomorphic and isometric with respect to $\|\cdot\|_{\mathbb{L}^2(\mathbb{R}^d)}$. Therefore, we can extend it to an isometric operator $\mathcal{F}_2 : \mathbb{L}^2(\mathbb{R}^d) \to \mathbb{L}^2(\mathbb{R}^d)$. This is captured in the following theorem whose proof can be found in Werner (2006).

Theorem 2.6.3 (Fourier transform of $\mathbb{L}^2(\mathbb{R}^d)$):
For $H \in \mathbb{L}^2(\mathbb{R}^d)$ and any $R \in \mathbb{R}_+$ we define the functions $F_R := \mathcal{F}[\mathbb{1}_{B_R(\mathbf{0})}H] \in \mathbb{L}^2(\mathbb{R}^d)$ where $B_R(\mathbf{0}) := \{\boldsymbol{x} \in \mathbb{R}^d : |\boldsymbol{x}|_{\mathbb{R}^d} \leq R\}$. Then $(F_R)_{R\in\mathbb{R}_+}$ converges in the $\mathbb{L}^2(\mathbb{R}^d)$ sense to the function $\mathcal{F}_2[H] \in \mathbb{L}^2(\mathbb{R}^d)$. Thus we define the PLANCHEREL-FOURIER TRANSFORM $\mathcal{F}_2[H] : \mathbb{R}^d \to \mathbb{C}$ of H by

$$\mathcal{F}_2[H](\boldsymbol{t}) := \lim_{R\to\infty} \frac{1}{(2\pi)^{d/2}} \int_{B_R(\mathbf{0})} H(\boldsymbol{x}) \exp(-i\langle \boldsymbol{x}, \boldsymbol{t}\rangle_{\mathbb{R}^d}) d\lambda^d(\boldsymbol{x})$$

for any $\boldsymbol{t} \in \mathbb{R}^d$ where the limit is defined in a $\|\cdot\|_{\mathbb{L}^2(\mathbb{R}^d)}$-sense.

Due to the construction the operator $\mathcal{F}_2 : \mathbb{L}^2(\mathbb{R}^d) \to \mathbb{L}^2(\mathbb{R}^d)$ it is an isometric isomorphism. The isometric property implies that the PLANCHEREL IDENTITY is valid for the transform $\mathcal{F}_2$, too. This is captured in the next Proposition.

Proposition 2.6.4 (Plancherel identity for the Fourier-Plancherel transform):
Let $H_1, H_2 \in \mathbb{L}^2(\mathbb{R}^d)$. Then $\langle H_1, H_2\rangle_{\mathbb{L}^2(\mathbb{R}^d)} = \langle \mathcal{F}_2[H_1], \mathcal{F}_2[H_2]\rangle_{\mathbb{L}^2(\mathbb{R}^d)}$. Further the PARSEVAL IDENTITY holds, that is $\|H_1\|_{\mathbb{L}^2(\mathbb{R}^d)} = \|\mathcal{F}_2[H_1]\|_{\mathbb{L}^2(\mathbb{R}^d)}$.

In fact, in the special case that $H \in \mathbb{L}^2(\mathbb{R}^d) \cap \mathbb{L}^1(\mathbb{R}^d)$, we can show that the two notions of Fourier transform coincide. The proof of the following Proposition can be found in Werner (2006).

Proposition 2.6.5:
For any $H \in \mathbb{L}^1(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$ holds true $\mathcal{F}_2[H] = \mathcal{F}[H]$ almost everywhere.

By an abuse of notation, we will drop the index 2 from now on. Furthermore we can deduce from Theorem 2.6.3 that there exists an inverse $\mathcal{F}_2^\dagger : \mathbb{L}^2(\mathbb{R}^d) \to \mathbb{L}^2(\mathbb{R}^d)$ of the Fourier-Plancherel transform which we will address now.

Theorem 2.6.6 (The inverse Fourier transform):
Let $H \in \mathbb{L}^1(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$. Then

$$\mathcal{F}^\dagger[H](\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} H(\boldsymbol{t}) \exp(i\langle \boldsymbol{t}, \boldsymbol{x}\rangle_{\mathbb{R}^d}) d\lambda^d(\boldsymbol{t})$$

for almost all $\boldsymbol{x} \in \mathbb{R}^d$.

Proof of Theorem 2.6.6. Let us define the shift operator $\mathrm{T} : \mathbb{L}^2(\mathbb{R}^d) \to \mathbb{L}^2(\mathbb{R}^d), H \mapsto T[H] : \boldsymbol{x} \mapsto H(-\boldsymbol{x})$. Then it is sufficient to show that $\mathrm{T} \circ \mathcal{F} = \mathcal{F}^\dagger$. Proposition 2.6.5 implies the statement. We first see that the inverse of the extension of $\mathcal{F} : \mathscr{S}(\mathbb{R}^d) \to \mathscr{S}(\mathbb{R}^d)$ equals the expansion of the inverse $\mathcal{F}^\dagger : \mathscr{S}(\mathbb{R}^d) \to \mathscr{S}(\mathbb{R}^d)$ since the last operator is continuous and has thus a unique expansion on $\mathbb{L}^2(\mathbb{R}^d)$. Now let $H \in \mathbb{L}^2(\mathbb{R}^d)$ be arbitrary and $(H_n)_{n\in\mathbb{N}}$ a sequence in $\mathscr{S}(\mathbb{R}^d)$ such that $\|H - H_n\|_{\mathbb{L}^2(\mathbb{R}^d)} \to 0$. Then due to the representation of $\mathcal{F}^\dagger : \mathbb{L}^2(\mathbb{R}^d) \to \mathbb{L}^2(\mathbb{R}^d)$ as the expansion of $\mathcal{F}^\dagger : \mathscr{S}(\mathbb{R}^d) \to \mathscr{S}(\mathbb{R}^d)$ we get $\mathcal{F}^\dagger[H] = \lim_{n\to\infty} \mathcal{F}^\dagger[H_n] = \lim_{n\to\infty}[T \circ \mathcal{F}][H_n]$ by Theorem 2.6.2. Now the statement follows from the continuity of $\mathcal{F}$ and T. □

2.6.2 Convolution

To begin this section, we see that for two measurable functions $H_1, H_2 : \mathbb{R}^d \to \mathbb{C}$ and $\boldsymbol{x} \in \mathbb{R}^d$ the function $I_{\boldsymbol{x}}; \mathbb{R}^d \to \mathbb{C}, \boldsymbol{y} \mapsto I_{\boldsymbol{x}}(\boldsymbol{y}) := H_1(\boldsymbol{x} - \boldsymbol{y})H_2(\boldsymbol{y})$ is measurable. This allows us to form the following, very general definition of the additive convolution of two functions.

Definition 2.6.7 (Additive convolution)**:**
Let $H_1, H_2 : \mathbb{R}^d \to \mathbb{C}$ be two measurable functions. Then for any $\boldsymbol{x} \in \mathbb{R}^d$ with $I_{\boldsymbol{x}} \in \mathbb{L}^1(\mathbb{R}^d)$ we define the ADDITIVE CONVOLUTION of H_1 and H_2 in the point $\boldsymbol{x}$ as

$$(H_1 *_+ H_2)(\boldsymbol{x}) := \int_{\mathbb{R}^d} H_1(\boldsymbol{x} - \boldsymbol{y}) H_2(\boldsymbol{y}) d\lambda^d(\boldsymbol{y}).$$

Due to the definition of the additive convolution, we can easily see that if $(H_1 *_+ H_2)(\boldsymbol{x})$ is well-defined, then $(H_1 *_+ H_2)(\boldsymbol{x}) = (H_2 *_+ H_1)(\boldsymbol{x})$. Further, if $p, q \in [1, \infty]$ are conjugate exponents, that is $p^{-1} + q^{-1} = 1$, then for all $H_1 \in \mathbb{L}^p(\mathbb{R}^d)$ and $H_2 \in \mathbb{L}^q(\mathbb{R}^d)$, $|(H_1 *_+ H_2)(\boldsymbol{x})| \leq \|H_1\|_{\mathbb{L}^p(\mathbb{R}^d)} \|H_2\|_{\mathbb{L}^q(\mathbb{R}^d)} < \infty$. In this scenario, the additive convolution of H_1 and H_2 is well-defined for any $\boldsymbol{x} \in \mathbb{R}^d$ which allows its interpretation as a bounded function $(H_1 *_+ H_2) : \mathbb{R}^d \to \mathbb{C}$. Another interpretation is given by the following proposition.

Proposition 2.6.8:
Let $H_1 \in \mathbb{L}^1(\mathbb{R}^d)$ and $H_2 \in \mathbb{L}^p(\mathbb{R}^d)$ for $p \in [1, \infty]$. Then $I_{\boldsymbol{x}}$ is for almost all $\boldsymbol{x} \in \mathbb{R}^d$ integrable and the (almost everywhere defined) function $(H_1 *_+ H_2) \in \mathbb{L}^p(\mathbb{R}^d)$ with $\|H_1 *_+ H_2\|_{\mathbb{L}^p(\mathbb{R}^d)} \leq \|H_1\|_{\mathbb{L}^1(\mathbb{R}^d)} \|H_2\|_{\mathbb{L}^p(\mathbb{R}^d)}$.

Proof of Proposition 2.6.8. The case $p = 1$ follows directly from the Fubini–Tonelli theorem. Let $p > 1$ and q its conjugate, that is $p^{-1} + q^{-1} = 1$. Then,

$$\begin{aligned} |H_1 *_+ H_2(\boldsymbol{y})| &\leq \int_{\mathbb{R}^d} |H_2(\boldsymbol{y})| |H_2(\boldsymbol{x} - \boldsymbol{y})|^{1/p} |H_2(\boldsymbol{x} - \boldsymbol{y})|^{1/q} d\lambda^d(\boldsymbol{x}) \\ &\leq \|H_1\|_{\mathbb{L}^1(\mathbb{R}^d)}^{1/q} \left(\int_{\mathbb{R}^d} |H_2(\boldsymbol{y})|^p |H_1(\boldsymbol{x} - \boldsymbol{y})| d\lambda^d(\boldsymbol{x}) \right)^{1/p} \end{aligned}$$

using the Hölder inequality. From this we deduce the claim after a series of elementary calculations. □

For the upcoming theory the case of $p = 1$ is the most interesting. In fact, if $H_1, H_2 \in \mathbb{L}^1(\mathbb{R}^d)$ then the additive convolution $H_1 *_+ H_2 \in \mathbb{L}^1(\mathbb{R}^d)$. This allows us to show the following statement known as the CONVOLUTION THEOREM.

Proposition 2.6.9 (Convolution theorem for the Fourier transform)**:**
Let $H_1, H_2 \in \mathbb{L}^1(\mathbb{R}^d)$. Then $\mathcal{F}[H_1 *_+ H_2] = \mathcal{F}[H_1] \cdot \mathcal{F}[H_2]$.

The proof of Proposition 2.6.9 can be done analogously to Proposition 2.2.5 and is thus omitted.

2.6.3 The Sobolev space

For a function $H \in \mathscr{S}(\mathbb{R}^d)$ and any multi index $\boldsymbol{\beta} \in \mathbb{N}_0^d$ we can show that $\mathcal{F}[D^{\boldsymbol{\beta}}[H]](\boldsymbol{t}) = (i\boldsymbol{t})^{\boldsymbol{\beta}} \mathcal{F}[H](\boldsymbol{t})$ for any $\boldsymbol{t} \in \mathbb{R}^d$, compare Werner (2006). To get a similar result for the equivalence classes of functions in $\mathbb{L}^2(\mathbb{R}^d)$ we will first introduce another notion of differentiability, the so-called WEAKLY DIFFERENTIABILITY.

Definition 2.6.10 (Weak differentiable, Sobolev spaces)**:**
Let $\boldsymbol{\beta} \in \mathbb{N}_0^d$ and $H \in \mathbb{L}^2(\mathbb{R}^d)$. Then we call $g \in \mathbb{L}^2(\mathbb{R}^d)$ the WEAK DERIVATIVE of H of order $\boldsymbol{\beta}$ if

$$\forall \varphi \in C_0^\infty(\mathbb{R}^d) : \langle g, \varphi \rangle_{\mathbb{L}^2(\mathbb{R}^d)} = (-1)^{|\boldsymbol{\beta}|_1} \langle H, \mathrm{D}^{\boldsymbol{\beta}}[\varphi] \rangle_{\mathbb{L}^2(\mathbb{R}^d)}.$$

If g exists, it is uniquely defined and we use the notation $\mathrm{D}^{(\boldsymbol{\beta})}[H] := g$. For $m \in \mathbb{N}_0$, we define the so-called SOBOLEV SPACE through

$$\mathcal{W}^m(\mathbb{R}^d) := \{H \in \mathbb{L}^2(\mathbb{R}^d) : \forall |\boldsymbol{\beta}|_1 \leq m : \mathrm{D}^{(\boldsymbol{\beta})}[H] \in \mathbb{L}^2(\mathbb{R}^d)\}.$$

The connection between weak differentiability and continuous differentiability is expressed by the following Theorem which is know as the LEMMA OF SOBOLEV.

Theorem 2.6.11 (Lemma of Sobolev)**:**
Let $m, k \in \mathbb{N}_0$ such that $m > k + d/2$. Then for any equivalence class $[H] \in \mathcal{W}^m(\mathbb{R}^d)$ there exist a representant of $[H]$ which is k-time continuously differentiable.

The proof of this theorem can be found in Werner (2006). Now let us come back to the interplay between the weak differentiability and the Fourier transform. In fact, we can proof the following property for weakly differentiable functions in $\mathbb{L}^2(\mathbb{R}^d)$ as done in Werner (2006).

Lemma 2.6.12:
Let $m \in \mathbb{N}_0$ and $H \in \mathcal{W}^m(\mathbb{R}^d)$. Then for any $\boldsymbol{\beta} \in \mathbb{N}_0^d$ with $|\boldsymbol{\beta}|_1 \leq m$ holds

$$\mathcal{F}[\mathrm{D}^{(\boldsymbol{\beta})} \circ H](\boldsymbol{t}) = (i\boldsymbol{t})^{\underline{\boldsymbol{\beta}}} \mathcal{F}[H](\boldsymbol{t}), \qquad \text{for any } \boldsymbol{t} \in \mathbb{R}^d.$$

The latter Lemma implies that if $H \in \mathcal{W}^m(\mathbb{R}^d)$, $t_j^m \mathcal{F}[H] \in \mathbb{L}^2(\mathbb{R}^d)$ for any $j \in [\![d]\!]$, where $t_j : \mathbb{R}^d \to \mathbb{R}, \boldsymbol{t} \mapsto t_j$ denotes the projection on the j-th component. After some basic calculations we can show that $(1 + |\boldsymbol{t}|^2)^{m/2} \mathcal{F}[H] \in \mathbb{L}^2(\mathbb{R}^d)$, which implies the relationship

$$\mathcal{W}^m(\mathbb{R}^d) \subseteq W^s(\mathbb{R}^d) := \{H \in \mathbb{L}^2(\mathbb{R}^d) : (1 + |\boldsymbol{t}|^2)^{s/2} \mathcal{F} \in \mathbb{L}^2(\mathbb{R}^d)\} \tag{2.5}$$

for any $s \leq m, s \in \mathbb{R}_+$. It is worth stressing out that in the case of $s = m$ we can show that $W^m(\mathbb{R}^d) \subseteq \mathcal{W}^m(\mathbb{R}^d)$ and thus the identity $W^m(\mathbb{R}^d) = \mathcal{W}^m(\mathbb{R}^d)$. It is therefore reasonable to speak of $W^s(\mathbb{R}^d)$ as an interpolation of $\mathcal{W}^m(\mathbb{R}^d)$ for non-integer parameter $s \in \mathbb{R}_+$. To distinguish between $\mathcal{W}^m(\mathbb{R}^d)$ and $W^s(\mathbb{R}^d)$, we will address the latter space as the FOURIER-SOBOLEV SPACE.

2.6.4 Anisotropic Sobolev spaces

In this subsection we briefly introduce the existing theory on anisotropic Sobolev spaces and anisotropic Fourier-Sobolev which is needed for a definition of equivalent spaces for the Mellin transform. The results can be found in Triebel (2006).

Definition 2.6.13 (Anisotropic Sobolev space)**:**
Let $\boldsymbol{m} = (m_1, \ldots, m_d)^T \in \mathbb{N}_0^d$. Then the ANISOTROPIC SOBOLEV SPACE is defined by

$$\mathcal{W}^{\boldsymbol{m}}(\mathbb{R}^d) := \{H \in \mathbb{L}^2(\mathbb{R}^d) : \forall j \in [\![d]\!] : \mathrm{D}^{(m_j \boldsymbol{e}_j)}[H] \in \mathbb{L}^2(\mathbb{R}^d)\}.$$

First we see that for any $\boldsymbol{m} \in \mathbb{N}_o$ with $m_1 = m_2 = \cdots = m_d = m \in \mathbb{N}_0$ we have $\mathcal{W}^{\boldsymbol{m}}(\mathbb{R}^d) = \mathcal{W}^m(\mathbb{R}^d)$. Additionally, we can show this second characterization of those spaces.

Proposition 2.6.14:
For $\boldsymbol{m} \in \mathbb{N}$

$$\mathcal{W}^{\boldsymbol{m}}(\mathbb{R}^d) = \{H \in \mathbb{L}^2(\mathbb{R}^d) : \sum_{\substack{\boldsymbol{\gamma} \in \mathbb{N}_0^d \\ |\boldsymbol{\gamma}/\boldsymbol{m}|_1 \leq 1}} \|\mathrm{D}^{(\boldsymbol{\gamma})}[H]\|^2_{\mathbb{L}^2(\mathbb{R}^d)} < \infty\}.$$

CHAPTER 3

Global density estimation under multiplicative measurement errors

Contents

In this chapter, we will use the results of Chapter 2 to study the estimation of a density function f of a positive random variable X, respectively random vector $\boldsymbol{X}$. We will use the Mellin transform to define an estimator of the density f and analyze its properties. The chapter is organized as follows.

We start in Section 3.1 by considering the estimation of the density $f : \mathbb{R}_+ \to \mathbb{R}_+$ using an independent, identically distributed sample $X_1, \dots, X_n$ drawn from f, which we will referred as the case of direct observations.

From there, we will consider in Section 3.2 the multiplicative measurement error model, that is, we assume that we have only have access to a sample $Y_1, \dots, Y_n$ of i.i.d. copies of $Y = XU$ where $X \sim f$ and U is a multiplicative, stochastic independent error term with known density g. Here, we will mainly focus on the case of so-called smooth error densities.

In Section 3.3 we will complete our study by including the case of super smooth error densities and discuss at which point the results differ from their analogues in the smooth error case.

Besides he smooth and super smooth error densities the class of oscillating error densities play a major role in Section 3.4. Here, our proposed estimator from Section 3.2, respectively Section 3.3, is not well-defined. We therefore introduce an estimator based on a ridge approach and study its properties.

In Section 3.5 we then leave the univariate case and see that our estimation strategy can be used for the anisotropic estimation of a multivariate density $f : \mathbb{R}_+^d \to \mathbb{R}_+$ in a multiplicative measurement

error model with smooth error density.
Returning to the univariate case, we consider in Section 3.6 the case of stationary $X_1, \ldots, X_n$ under multiplicative measurement errors with a digression to stochastic volatility models.

3.1 The case of direct observation

3.1.1 Introduction

In this section, we are interested in estimating the unknown density $f : \mathbb{R}_+ \to \mathbb{R}_+$ of a positive random variable X given independent and identically distributed (i.i.d.) copies of X.
In the literature, non-parametric density estimation is a well-discussed problem and many estimation strategies based on splines, kernels or wavelets, to name but a few, are considered. For an overview of various methods we refer to Comte (2017), Efromovich (1999), Silverman (2018) and Tsybakov (2009).
In this work, we consider a density estimator using the Mellin transform and a spectral cut-off regularization of its inverse, which borrows ideas from Belomestny and Goldenshluger (2020) where they considered the model of multiplicative measurement errors which will be considered in Section 3.2. In this section, we will restrict ourselves to the case that the direct observations are given. Our estimation strategy differs in the following way from the existing results for density estimation on $\mathbb{R}_+$. We define a density estimator by making use of the the Mellin transform and applying an additional spectral cut-off on the inversion of the Mellin-transform. We measure the accuracy of the estimator by introducing a global risk in terms of a weighted $\mathbb{L}^2$-norm. Exploiting properties of the Mellin transform we characterize the natural regularity conditions which borrow ideas from the inverse problems community (Engl et al. (1996)). The proposed estimator, however, involves a tuning parameter which is selected by a data-driven method. We establish an oracle inequality for the fully-data driven spectral cut-off estimator.
Moreover we show that uniformly over Mellin-Sobolev spaces the proposed data-driven estimator is minimax-optimal. We state both an upper bound for the mean weighted integrated squared error of the fully-data driven spectral cut-off estimator and a general lower bound for estimating the density f based on i.i.d. copies from f.
This section is structured as follows. In Subsection 3.1.2 we use the properties of the Mellin transform to introduce and analyze our density estimator based on direct observations $X_1, \dots, X_n$ drawn from f. We derive an oracle type upper bound for its mean weighted integrated squared error. In Subsection 3.1.3 we show that the presented estimator is minimax optimal over Mellin-Sobolev spaces. A fully data-driven estimator is presented and analyzed in Subsection 3.1.4. Finally, results of a simulation study are reported in Subsection 3.1.5 which visualize the reasonable finite sample performance of our estimators. Proofs of the results of the Subsection 3.1.2, 3.1.3 and 3.1.4 can be found in the Subsection 3.1.7.

3.1.2 Estimation strategy

Revisiting the approximation study in Section 2.5 we defined an approximation in Equation (2.4) of an arbitrary function $h \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ by

$$h_k(x) := \mathcal{M}_c^\dagger[\mathcal{M}_c[h]\mathbb{1}_{[-k,k]}](x) = \frac{1}{2\pi}\int_{[-k,k]} x^{-c-it}\mathcal{M}_c[h](t)d\lambda(t), \tag{3.1}$$

for any $x, k \in \mathbb{R}_+$, and analyze this approximation theoretically by considering Mellin-Sobolev spaces and practically via a numerical study.
In this subsection, we use this approximation step for the unknown density $f \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ to define the family of approximations $(f_k)_{k\in\mathbb{R}_+}$ using Equation (3.1) for $h = f$. In a second step, we

will substitute each unknown quantity by an empirical counterpart to define a family of unbiased estimators $(\widehat{f}_k)_{k\in\mathbb{R}_+}$ of $(f_k)_{k\in\mathbb{R}_+}$. We will then close this subsection, by giving an upper bound for the mean integrated weighted squared error of the estimator $\widehat{f}_k$, that is $\mathbb{E}^n_f(\|f-\widehat{f}_k\|^2_{x^{2c-1}})$ and by showing that we can find a sequence $(k_n)_{n\in\mathbb{N}}$ such that $\mathbb{E}^n_f(\|f-\widehat{f}_{k_n}\|^2_{x^{2c-1}})$ vanishes for $n\to\infty$. Here, $\mathbb{E}^n_f$ denotes the expectation with respect to the distribution $\mathbb{P}^n_f$ of $(X_1,\dots,X_n)$. For the case $n=1$ we set $\mathbb{E}_f=\mathbb{E}^1_f$, respectively $\mathbb{P}_f=\mathbb{P}^1_f$.

The empirical Mellin transform Considering Equation (3.1) we see that in order to define an unbiased estimator for the function f_k, $k\in\mathbb{R}_+$, we are in need of an unbiased estimator of $\mathcal{M}_c[f](t)$ for any $t\in[-k,k]$. Keeping in mind that for a density f and any $t\in\mathbb{R}$ we have $\mathcal{M}_c[f](t)=\mathbb{E}_f(X^{c-1+it})$, a natural unbiased estimation strategy is given by a moment estimator, that is

$$\widehat{\mathcal{M}}_c(t):=n^{-1}\sum_{j=1}^n X_j^{c-1+it}\quad\text{for any } t\in\mathbb{R},$$

which we call the EMPIRICAL MELLIN TRANSFORM of the sample $X_1,\dots,X_n$. It is worth stressing out, that for the empirical distribution $\widehat{\mathbb{P}}_n:=n^{-1}\sum_{j=1}^n\delta_{X_j}$ we have that $\mathcal{M}_c[\widehat{\mathbb{P}}_n]=\widehat{\mathcal{M}}_c$ where the definition of the Mellin transform of a distribution is given in Definition 2.1.5. Here δ_x denotes the Dirac measure in the point x.
Considering the function $\widehat{\mathcal{M}}_c:\mathbb{R}\to\mathbb{C}$, we see that $|\widehat{\mathcal{M}}_c(t)|\le\widehat{\mathcal{M}}_c(0)<\infty$ almost surely, which implies that $\mathbb{1}_{[-k,k]}\widehat{\mathcal{M}}_c\in\mathbb{L}^1(\mathbb{R})\cap\mathbb{L}^2(\mathbb{R})$. Applying now the inversion formula, Proposition 2.3.6, the functions $\widehat{f}_k:=\mathcal{M}_c^\dagger[\mathbb{1}_{[-k,k]}\widehat{\mathcal{M}}_c]$ lies in $\mathbb{L}^2(\mathbb{R}_+,x^{2c-1})$ and can be expressed explicitly by

$$\widehat{f}_k(x)=\frac{1}{2\pi}\int_{[-k,k]}x^{-c-it}\widehat{\mathcal{M}}_c(t)d\lambda(t)\quad\text{for any } x\in\mathbb{R}_+.\tag{3.2}$$

Other estimators Let us shortly motivate the estimator given in Equation (3.2) from another point of view. Considering the inverse formula for Mellin transform, Proposition 2.3.6, we get $f=\mathcal{M}_c^\dagger[\mathcal{M}_c[f]]$. A naive approach here, would be to plug the empirical Mellin transform in directly to define an estimator $\widehat{f}=\mathcal{M}_c^\dagger[\widehat{\mathcal{M}}_c]$. This is not well-defined since $\widehat{\mathcal{M}}_c$ is not square-integrable and does thus not lay in the domain of $\mathcal{M}_c^\dagger$. In the inverse problems literature, these situation are typically dealt with using regularized version of the operator $\mathcal{M}_c^\dagger$ such that $\widehat{\mathcal{M}}_c$ lies in it domain. One of the most frequently used techniques for regularizing an operator is the so-called spectral cut-off method, compare Engl et al. (1996). Applying this method to the theoretical estimator $\widehat{f}=\mathcal{M}_c^\dagger[\widehat{\mathcal{M}}_c]$ leads to the estimator in Equation (3.2).
Another often used method, to construct a well-defined estimator from $\widehat{f}=\mathcal{M}_c^\dagger[\widehat{\mathcal{M}}_c]$, would be a kernel estimator. Let us consider a function $\kappa\in\mathbb{L}^2(\mathbb{R})\cap\mathbb{L}^1(\mathbb{R})$ and define $\kappa_h(x):=\kappa(hx), x\in\mathbb{R}$ and $h\in\mathbb{R}_+$. Then due to the boundness of $\widehat{\mathcal{M}}_c$ we may define for any $h\in\mathbb{R}_+$ the estimator $\widehat{f}_h=\mathcal{M}_c^\dagger[\kappa_h\widehat{\mathcal{M}}_c]$ which is explicitly expressed,

$$\widehat{f}_h(x)=\frac{1}{2\pi}\int_{\mathbb{R}}x^{-c-it}\kappa(ht)\widehat{\mathcal{M}}_c(t)d\lambda(t)\quad\text{for any } x\in\mathbb{R}_+.\tag{3.3}$$

Furthermore, defining $K_h := \mathcal{M}_c^\dagger[\kappa_h] \in \mathbb{L}^2(\mathbb{R})$ we can rewrite the estimator $\widehat{f}_h$ and rearrange it in the following form

$$\widehat{f}_h(x) = \frac{1}{n}\sum_{j=1}^n X_j^{-1}\frac{1}{2\pi}\int_{\mathbb{R}}\left(\frac{x}{X_j}\right)^{-c-it}\kappa(ht)d\lambda(t) = \frac{1}{n}\sum_{j=1}^n X_j^{-1}K_h(x/X_j), \quad \text{for any } x\in\mathbb{R}_+.$$

For the special case of $\kappa = \mathbb{1}_{[-1,1]}$ and $h = 1/k$ the estimators $\widehat{f}_h$ in Equation (3.3) and $\widehat{f}_k$ in Equation (3.2) coincide. In this case, the function K_h would be given by

$$K_h(x) = x^{-c}\frac{1}{2\pi}\int_{[-h^{-1},h^{-1}]}\exp(-it\log(x))d\lambda(t) = \frac{x^{-c}}{\log(x)}\sin(\log(x)h^{-1}) \quad \text{for any } x\in\mathbb{R}_+,$$

exploiting that $\sin(y) = (\exp(iy)-\exp(-iy))/(2\pi i)$ for any $y\in\mathbb{R}$. This kernel resembles strongly the sinc-kernel which is used frequently for additive deconvolution problems. For this dissertation we will mainly focus on the spectral cut-off estimator given in Equation (3.2). The point-wise estimation using kernel estimators based on Mellin transforms is intensively studied in Belomestny and Goldenshluger (2020).

Upper bound for the risk Now let us consider the risk $\mathbb{E}_f^n(\|f-\widehat{f}_k\|_{x^{2c-1}}^2)$ for $k\in\mathbb{R}_+$ and $\widehat{f}_k$ defined in Equation (3.2). By construction of the function $\widehat{f}_k$ we can ensure that $\widehat{f}_k \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ almost-surely. This allows us to apply the Parseval identity, Proposition 2.3.5, in the following way

$$\|f-\widehat{f}_k\|_{x^{2c-1}}^2 = \frac{1}{2\pi}\|\mathcal{M}_c[f]-\mathcal{M}_c[\widehat{f}_k]\|_{\mathbb{R}}^2 = \frac{1}{2\pi}\left(\|\mathcal{M}_c[f-f_k]\|_{\mathbb{R}}^2 + \|\mathcal{M}_c[f_k-\widehat{f}_k]\|_{\mathbb{R}}^2\right)$$

using that $\mathcal{M}_c[f-f_k]$ and $\mathcal{M}_c[f_k-\widehat{f}_k]$ have disjoint supports. Again, by applying the Parseval identity, we see that the first term is the approximation error term considered in Section 2.5. In nonparametric Statistics this term is usually referred as the SQUARED BIAS-TERM.
Considering the second term, it is dependent on the estimator $\widehat{f}_k$ and thus random. Though it is almost-surely finite, we have not given any sufficient condition to ensure that it possesses a finite first moment. This can be done by a simple moment condition. In fact, assuming that $\mathbb{E}_f(X_1^{2(c-1)}) < \infty$, we can state that $\mathbb{E}_f^n(|\widehat{\mathcal{M}}_c(t)-\mathcal{M}_c(t)|^2) = n^{-1}\mathbb{V}\mathrm{ar}_f(X_1^{c-1+it}) \le n^{-1}\mathbb{E}_f(X_1^{2(c-1)})$ for all $t\in\mathbb{R}$. This allows us to use the Fubini-Tonelli theorem and get

$$\frac{1}{2\pi}\mathbb{E}_f^n(\|\mathcal{M}_c[f_k-\widehat{f}_k]\|_{\mathbb{R}}^2) = \frac{1}{2\pi}\int_{[-k,k]}\mathbb{E}_f^n(|\widehat{\mathcal{M}}_c(t)-\mathcal{M}_c(t)|^2)d\lambda(t) \le \frac{k}{n\pi}\mathbb{E}_f(X_1^{2(c-1)}).$$

The term $(2\pi)^{-1}\mathbb{E}_f^n(\|\mathcal{M}_c[f_k-\widehat{f}_k]\|_{\mathbb{R}}^2) = \mathbb{E}_f^n(\|f_k-\widehat{f}_k\|_{x^{2c-1}}^2)$ is called the VARIANCE-TERM and the decomposition of the risk $\mathbb{E}_f^n(\|f-\widehat{f}_k\|_{x^{2c-1}}^2)$ into the squared bias term $\|f-f_k\|_{x^{2c-1}}^2$ and the variance term $\mathbb{E}_f^n(\|f_k-\widehat{f}_k\|_{x^{2c-1}}^2)$ is a frequently used tool in non-parametric statistics called the BIAS-VARIANCE DECOMPOSITION. We collect these consideration in the following Proposition which we have already proven.

Proposition 3.1.1 (Upper bound of the risk)**:**
Let $f\in\mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ and $\sigma_c := \mathbb{E}_f(X_1^{2(c-1)}) < \infty$. Then for all $k\in\mathbb{R}_+$,

$$\mathbb{E}_f^n(\|f-\widehat{f}_k\|_{x^{2c-1}}^2) = \|f-f_k\|_{x^{2c-1}}^2 + \mathbb{E}_f^n(\|f_k-\widehat{f}_k\|_{x^{2c-1}}^2) \le \|f-f_k\|_{x^{2c-1}}^2 + \frac{k}{n\pi}\sigma_c$$

where $\widehat{f}_k$ is the estimator defined in Equation (3.2).

Let us have a closer look at the squared bias and the variance term. While the squared bias term is decreasing and tending to 0 for increasing $k \in \mathbb{R}_+$, the variance term is increasing. To ensure that both term decrease in $n \in \mathbb{N}$, it is sufficient to choose a sequence $(k_n)_{n\in\mathbb{N}}$ of spectral cut-off parameters with $k_n \to \infty$ and $k_n n^{-1} \to 0$ for $n \to \infty$. This thought implies directly the following Corollary whose proof is thus omitted.

Corollary 3.1.2 (Consistency)**:**
Let $f \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ and $\sigma_c = \mathbb{E}_f(X_1^{2c-1}) < \infty$. Then for every sequence $(k_n)_{n\in\mathbb{N}}$ with $k_n \to \infty$ and $k_n n^{-1} \to 0$ for $n \to \infty$,

$$\mathbb{E}_f^n(\|f - \widehat{f}_{k_n}\|_{x^{2c-1}}^2) \to 0,$$

implying that $\|f - \widehat{f}_{k_n}\|_{x^{2c-1}}^2 \xrightarrow{\mathbb{P}} 0$.

Although the simple assumption on $(k_n)_{n\in\mathbb{N}}$ leads to a consistent estimator only based on the data $X_1, \dots, X_n$, we might ask how different choices of a sequence $(k_n)_{n\in\mathbb{N}}$ might effect the rate with which the risk is decaying. Comparing to Section 2.5, we have seen that the underlying density is strongly connected to the decay of the bias and thus affects the rate of the risk. A more sophisticated analysis of the bias term and the corresponding risk, will be done in the following subsection.

3.1.3 Minimax theory

In this subsection, we will build ellipsoids in $\mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ based on the Mellin-Sobolev spaces introduced in Definition 2.3.9 and deduce a rate for the upper bound of the risk uniformly over the ellipsoids. This gives us the means to derive an upper bound for the minimax risk of the density estimation over the ellipsoids. We will prove that the stated rate is minimax by presenting a lower bound proof. We first state both the upper and lower bound for the risk. Afterwards, we interpret these results in the context of minimax theory.

Definition 3.1.3 (Mellin Sobolev ellipsoids)**:**
For $s, L \in \mathbb{R}_+$ and $c \in \mathbb{R}$ we define the MELLIN-SOBOLEV ELLIPSOIDS by

$$\mathbb{W}_c^s(L) := \{f \in \mathbb{W}_c^s(\mathbb{R}_+) : |f|_{s,c}^2 := \|(1+t^2)^{s/2}\mathcal{M}_c[f]\|_{\mathbb{R}}^2 \leq L\}.$$

Then for $f \in \mathbb{W}_c^s(L)$ holds

$$\|f - f_k\|_{x^{2c-1}}^2 = \frac{1}{\pi}\int_k^\infty |\mathcal{M}_c[f](t)|^2(1+t^2)^s(1+t^2)^{-s}d\lambda(t) \leq k^{-2s}\frac{L}{2\pi}.$$

Furthermore, because f is a density, and in need to control the variance term, it is natural to define the following ellipsoids

$$\mathbb{D}_c^s(L) := \{f \in \mathbb{W}_c^s(L) : f \text{ is a density }, \sigma_c = \mathbb{E}_f(X_1^{2(c-1)}) \leq L\}.$$

Then we can state for $f \in \mathbb{D}_c^s(L)$ that $\mathbb{E}_f^n(\|f_k - \widehat{f}_k\|_{x^{2c-1}}^2) \leq \frac{Lk}{\pi n}$, which proves the following Proposition.

Proposition 3.1.4 (Upper bound for the minimax risk)**:**
Let $s, L \in \mathbb{R}_+$ and $c \in \mathbb{R}$. Then for $k_o := n^{1/(2s+1)}$ we get

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_f^n(\|f - \widehat{f}_{k_o}\|_{x^{2c-1}}^2) \leq C(L, s) n^{-2s/(2s+1)}$$

where $C(L) := 3L/(2\pi)$ is a positive constant depending on L.

Despite the fact, that this rate is typical for nonparametric estimation of densities, compare Comte (2017) or Tsybakov (2009), we might ask ourselves if there exists an estimator based on the sample $X_1, \ldots, X_n$ which can access a better rate uniformly over the ellipsoids $\mathbb{D}_c^s(L)$. The answer to this question is captured in the following Theorem.

Theorem 3.1.5 (Lower bound for the minimax risk)**:**
Let $s \in \mathbb{N}$ and $c \in \mathbb{R}$. Then there exist constants $L_{s,c}, n_s > 0$ such that for all $L \geq L_{s,c}$, $n \geq n_s$ and for any estimator $\widehat{f}$ of f based on an i.i.d. sample $(X_j)_{j \in [\![n]\!]}$,

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_f^n(\|\widehat{f} - f\|_{x^{2c-1}}^2) \geq C_c n^{-2s/(2s+1)}.$$

where C_c is a positive constant depending on c.

Proof of Theorem 3.1.5. First we outline the main steps of the proof. The upcoming Lemmata 3.1.6–3.1.8 will be proven afterwards. We will construct a family of functions in $\mathbb{D}_c^s(L)$ by a perturbation of the density $f_o : \mathbb{R}_+ \to \mathbb{R}_+$ with small bumps, such that their $\mathbb{L}^2(\mathbb{R}, x^{2c-1})$-distance and the Kullback-Leibler divergence of their induced distributions can be bounded from below and above, respectively. The claim follows by applying Theorem 2.5 in Tsybakov (2009). We use the following construction.
Let $\psi \in C_0^\infty(\mathbb{R})$ be a function with support in $[0, 1]$ and $\int_{(0,1)} \psi(x) d\lambda(x) = 0$. For each $K \in \mathbb{N}$ (to be selected below) and $k \in [\![0, K-1]\!]$ we define the bump-functions $\psi_{k,K}(x) := \psi(xK - K - k)$, $x \in \mathbb{R}$ and define for $j \in \mathbb{N}_0$ the finite constant $C_{j,\infty} := \max(\|\psi^{(l)}\|_\infty, l \in [\![0, j]\!])$. For a bump-amplitude $\delta > 0$ and a vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K) \in \{0, 1\}^K$ we define

$$f_{\boldsymbol{\theta}}(x) = f_o(x) + \delta K^{-s} \sum_{k=0}^{K-1} \theta_{k+1} \psi_{k,K}(x), \tag{3.4}$$

where $f_o : \mathbb{R}_+ \to \mathbb{R}_+, x \mapsto \frac{1}{\sqrt{2\pi x}} \exp(-\log(x)^2/2) \mathbb{1}_{\mathbb{R}_+}(x)$ is the density of a Log-Normal distribution, compare Example 2.1.4. Until now, we did not give a sufficient condition to ensure that our constructed functions $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \{0, 1\}^K\}$ are in fact densities. This condition is given by the following lemma.

Lemma 3.1.6:
Let $0 < \delta < \delta_o(\psi) := f_o(2)/C_{0,\gamma}$. Then for all $\boldsymbol{\theta} \in \{0, 1\}^K$, $f_{\boldsymbol{\theta}}$ is a density.

Further, we can show that these densities all lie inside the ellipsoids for L big enough. This is captured in the following lemma.

Lemma 3.1.7:
Let $s \in \mathbb{N}$. Then, there is $L_{s,c,\delta} > 0$ such that f_o and any $f_{\boldsymbol{\theta}}$ as in (3.4) with $\boldsymbol{\theta} \in \{0, 1\}^K$, $K \in \mathbb{N}$, belong to $\mathbb{D}_c^s(L)$ for all $L \geq L_{s,c,\delta}$.

Exploiting Varshamov-Gilbert's lemma (see Tsybakov (2009)) in the proof of Lemma 3.1.8 we show further that there is $M \in \mathbb{N}$ with $M \geq 2^{K/8}$ and a subset $\{\boldsymbol{\theta}^{(0)}, \dots, \boldsymbol{\theta}^{(M)}\}$ of $\{0,1\}^K$ with $\boldsymbol{\theta}^{(0)} = (0, \dots, 0)$ such that for all $j, l \in [\![0, M]\!]$, $j \neq l$ the $\mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$-distance and the Kullback-Leibler divergence are bounded for $K \geq 8$. Here, we define for two Lebesgue densities f_1, f_0 on $\mathbb{R}_+$ its Kullback-Leibler divergence by

$$\mathrm{KL}(f_1, f_0) := \begin{cases} \int_{\mathbb{R}} \log(\frac{d\mathbb{P}_{f_1}}{d\mathbb{P}_{f_0}}(x)) d\mathbb{P}_{f_0}(x) & \mathbb{P}_{f_1} \ll \mathbb{P}_{f_0} \\ \infty, & \text{else.} \end{cases}$$
$$= \begin{cases} \int_{\mathbb{R}_+} \log(\frac{f_1(x)}{f_o(x)}) f_o(x) d\lambda(x), & \mathbb{P}_{f_1} \ll \mathbb{P}_{f_0} \\ \infty, & \text{else.} \end{cases}$$

Lemma 3.1.8:
Let $K \geq 8$. Then there exists a subset $\{\boldsymbol{\theta}^{(0)}, \dots, \boldsymbol{\theta}^{(M)}\}$ of $\{0,1\}^K$ with $\boldsymbol{\theta}^{(0)} = (0, \dots, 0)$ such that $M \geq 2^{K/8}$ and for all $j, l \in [\![0, M]\!]$, $j \neq l$,

$$\begin{aligned} &(i) \quad \|f_{\boldsymbol{\theta}^{(j)}} - f_{\boldsymbol{\theta}^{(l)}}\|^2_{x^{2c-1}} \geq C_c \|\psi\|^2_{x^0} \delta^2 K^{-2s} \\ &(ii) \quad \mathrm{KL}(f_{\boldsymbol{\theta}^{(j)}}, f_{\boldsymbol{\theta}^{(0)}}) \leq C_{f_o} \|\psi\|^2_{x^0} \delta^2 \log(M) K^{-2s-1} \end{aligned}$$

and where KL is the Kullback-Leibler-divergence.

Selecting $K = \lceil n^{1/(2s+1)} \rceil$,

$$\frac{1}{M} \sum_{j=1}^{M} \mathrm{KL}((f_{\boldsymbol{\theta}^{(j)}})^{\otimes n}, (f_{\boldsymbol{\theta}^{(0)}})^{\otimes n}) = \frac{n}{M} \sum_{j=1}^{M} \mathrm{KL}(f_{\boldsymbol{\theta}^{(j)}}, f_{\boldsymbol{\theta}^{(0)}}) \leq c_{\psi, f_o} \delta^2 \log(M).$$

Thus, there exists a $\delta(\psi, f_o) \in \mathbb{R}_+$ such that $c_{\psi, f_o}\delta^2 < 1/8$ for all $\delta \leq \delta_1(\psi, f_o)$ and $M \geq 2$ for $n \geq n_s := 8^{2s+1}$. Thereby, we can use Theorem 2.5 of Tsybakov (2009), which in turn for any estimator $\widehat{f}$ of f implies

$$\sup_{f \in \mathbb{D}^s_c(L)} \mathbb{P}^n_f(\|\widehat{f} - f\|^2_{x^{2c-1}} \geq \frac{c_{\psi,\delta,c}}{2} n^{-2s/(2s+1)}) \geq \frac{\sqrt{M}}{1+\sqrt{M}} (1 - 1/4 - \sqrt{\frac{1}{4\log(M)}}) \geq 0.07.$$

Note that the constant $c_{\psi,\delta,c}$ only depends on ψ, c and δ, hence it is independent of the parameters s, L and n. The claim of Theorem 3.1.5 follows by using Markov's inequality, which completes the proof. □

Now that we have shown all necessary theorems we will interpret our results in the context of the minimax theory.

Definition 3.1.9 (Minimax Risk):
The minimax risk of the density estimation over the ellipsoids $\mathbb{D}^s_c(L)$ given the i.i.d. sample $X_1, \dots, X_n$ is defined as

$$\mathcal{R}_n := \inf_{\widehat{f}} \sup_{f \in \mathbb{D}^s_c(L)} \mathbb{E}^n_f(\|\widehat{f} - f\|^2_{x^{2c-1}}) \tag{3.5}$$

where the infimum is taken over all estimators based on the sample $X_1, \dots, X_n$.

Of special interest in this context is the determination of the so-called MINIMAX RATE $(r_n)_{n\in\mathbb{N}}$ which is defined as follows.

Definition 3.1.10 (Minimax rate)**:**
Let $(\mathcal{R}_n)_{n\in\mathbb{N}}$ be the minimax risk. Then a sequence $(r_n)_{n\in\mathbb{N}}$ is called a MINIMAX RATE if there exists constants $C_1, C_2 > 0$ such that for all $n \in \mathbb{N}$,

$$C_1 r_n \leq \mathcal{R}_n \leq C_2 r_n.$$

Inspired by Theorem 3.1.5, we will now show that $(r_n)_{n\in\mathbb{N}}$ is given by $r_n = n^{-2s/(2s+1)}$. While the lower bound is covered in Theorem 3.1.5, the upper bound is achieved by the simple idea that for any fixed estimator $\widehat{f}_o$ based on $X_1, \dots, X_n$,

$$\mathcal{R}_n \leq \sup_{f\in\mathbb{D}_c^s(L)} \mathbb{E}_f^n(\|\widehat{f}_o - f\|^2_{x^{2c-1}})$$

In other words, to show that a given lower bound is optimal, it is sufficient to show that there exists an estimator which can indeed achieve the proposed rate. Now, Proposition 3.1.4 implies that our estimator $\widehat{f}_{k_{or}}$ achieves the desired rate which implies that $(r_n)_{n\in\mathbb{N}}$ is in fact, the minimax rate. Furthermore, we call an estimator which achieves the minimax rate MINIMAX OPTIMAL. This is captured in the following proposition.

Proposition 3.1.11:
Let $c \in \mathbb{R}$ and $s \in \mathbb{N}$. Then for any $L \geq L_{s,c}$ (given in Theorem 3.1.5) and $n \in \mathbb{N}$ we can state that the estimator $\widehat{f}_{k_{or}}$ with $k_{or} = n^{1/(2s+1)}$ is minimax-optimal with minimax rate $n^{-2s/(2s+1)}$.

Although the choice of k_{or} is independent on the density f itself, it is still dependent on the regularity parameter $s \in \mathbb{R}_+$. It might be tempting to choose $s \in \mathbb{R}_+$ fixed and interpret it as a regularity assumption, but this could cause a massive error due to the fact that the choice k_{or} is rather pessimistic which will be explained in the next paragraph.

Faster rates Revisiting the examples of the approximations and numerical study, Section 2.5, we have seen that the family of Gamma-distributions and Log-Normal distributions possess Mellin transforms which decay exponentially. As a short reminder the density and its Mellin transform of a Gamma-distribution with parameters $p, \lambda \in \mathbb{R}_+$ is given by

$$f(x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} \exp(-\lambda x)\mathbb{1}_{\mathbb{R}_+}(x) \text{ and } \mathcal{M}_c[f](t) = \lambda^{1-c-it}\frac{\Gamma(p+c-1+it)}{\Gamma(p)}$$

for any $x, \lambda \in \mathbb{R}_+$, $t \in \mathbb{R}$ and $c > -p$. For the Log-Normal distribution we have for the parameter $\mu \in \mathbb{R}$ and $\lambda^2 \in \mathbb{R}_+$,

$$f(x) = \frac{1}{\sqrt{2\pi\lambda^2}x} \exp\left(-\frac{1}{2\lambda^2}(\log(x)-\mu)^2\right) \text{ and } \mathcal{M}_c[f](t) = e^{\mu(c-1+it)} \exp\left(\frac{\lambda^2(c-1+it)^2}{2}\right)$$

for any $x \in \mathbb{R}_+$ and $t, c \in \mathbb{R}$. Applying for the Gamma distribution the Stirling formula in Andrews et al. (1999), which states that $|\Gamma(\sigma + it)| \leq |t|^{\sigma-1/2}\exp(-\pi|t|/2)$ for $|t| \geq 2$ and $\sigma \geq -2$, we get for both cases that

$$\|f - f_k\|^2_{x^{2c-1}} = \frac{1}{\pi}\int_k^\infty |\mathcal{M}_c[f](t)|^2 d\lambda(t) \leq C_f \exp(-\alpha k^a)$$

for $\alpha \in \mathbb{R}_+$, dependent on f and $a = 1$ for the Gamma distribution and $a = 2$ for the Log-Normal distribution. Now choosing $k_n = \log(n)^{1/a}/\alpha^{-1}$ we can ensure that for both cases

$$\inf_{k \in \mathbb{R}_+} \left(\|f - f_k\|^2_{x^{2c-1}} + \mathbb{E}^n_f(\|\widehat{f}_k - f_k\|^2_{x^{2c-1}}) \right) \leq \mathbb{E}^n_f(\|\widehat{f}_{k_n} - f\|^2_{x^{2c-1}}) \leq C_f \frac{\log(n)^{1/a}}{n}, \tag{3.6}$$

leading to a faster rate than provided by Proposition 3.1.4 for any choice of $s \in \mathbb{R}_+$. In other words, if we interpret $s \in \mathbb{R}_+$ as a fixed regularity assumption, and choosing $k_n = n^{1/(2s+1)}$ we may end up in many cases with a slower rate. It is therefore desirable, to find a fully data-driven choice $\widehat{k}$, only dependent on $(X_j)_{j \in [\![n]\!]}$ such that the risk of $\widehat{f}_{\widehat{k}}$ behaves up to a negligible term like the infimum on the left hand side of Equation (3.6). In the next subsection, we consider a choice of the parameter $k \in \mathbb{R}_+$ based only on the observation $X_1, \dots, X_n$ which can achieve even the faster rate presented in (3.6) without any prior knowledge about the decay of the bias.

3.1.4 Data-driven method

We will now introduce a data-driven choice of the cut-off parameter $k \in \mathbb{R}_+$. To do so, we first define first a contrast for a nested system of subspaces $(L_k)_{k \in \mathbb{R}_+}$ of $\mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$, that is $L_k \subseteq L_{k'}$ for all $k \leq k' \in \mathbb{R}_+$ and show that the estimator presented in Equation (3.2) can be interpreted as the corresponding minimal contrast estimator. In this situation, we can then use a penalized contrast approach, called model selection, to define our data-driven estimator and present an upper bound of its risk.
Let us define the nested system of subspaces $(L_k)_{k \in \mathbb{R}_+}$ by

$$L_k := \{h \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) : \operatorname{supp}(\mathcal{M}_c[h]) := \{t \in \mathbb{R} : \mathcal{M}_c[h](t) \neq 0\} \subseteq [-k, k]\}.$$

Obviously, we see that $L_k \subseteq L_{k'}$ for any $k \leq k'$. For any $k \in \mathbb{R}_+$ and $h \in L_k$ we define the empirical contrast $\gamma_n(h)$ by

$$\gamma_n(h) := \|h\|^2_{x^{2c-1}} - 2\frac{1}{2\pi} \int_{\mathbb{R}} \widehat{\mathcal{M}}_c(t) \mathcal{M}_c[h](-t) d\lambda(t) \tag{3.7}$$

which is well-defined since $|\widehat{\mathcal{M}}_c(t)| \leq \widehat{\mathcal{M}}_c(0)$ is bounded and $\mathcal{M}_c[h]$ has compact support. In expectation we deduce from (3.7),

$$\begin{aligned}\mathbb{E}^n_f(\gamma_n(h)) &= \|h\|^2_{x^{2c-1}} - 2\frac{1}{2\pi} \int_{\mathbb{R}} \mathcal{M}_c[f](t) \mathcal{M}_c[h](-t) d\lambda(t) \\ &= \|h\|^2_{x^{2c-1}} - 2\langle h, f \rangle_{x^{2c-1}} = \|h - f\|^2_{x^{2c-1}} - \|f\|^2_{x^{2c-1}}\end{aligned}$$

by applying the Fubini-Tonelli theorem and the Plancherel identity given in Proposition 2.3.5. As far as the contrast $\gamma(h) := \mathbb{E}^n_f(\gamma_n(h))$ for $h \in L_k$ is converned, we can easily see that

$$f_k = \arg\min_{h \in L_k} \gamma(h), \quad \text{ since } \langle f, h \rangle_{x^{2c-1}} = \langle f_k, h \rangle_{x^{2c-1}}$$

for any $h \in L_k$. Analogously, we have

$$\frac{1}{2\pi} \int_{\mathbb{R}} \widehat{\mathcal{M}}_c(t) \mathcal{M}_c[h](t) d\lambda(t) = \frac{1}{2\pi} \langle \widehat{\mathcal{M}}_c \mathbb{1}_{[-k,k]}, h \rangle_{\mathbb{R}} = \langle \widehat{f}_k, h \rangle_{x^{2c-1}}$$

implying that we can write $\gamma_n(h) = \|\widehat{f}_k - h\|^2_{x^{2c-1}} - \|\widehat{f}_k\|^2_{x^{2c-1}}$, which is obviously minimal for $h = \widehat{f}_k$.

Proposition 3.1.12 (Minimal contrast estimator)**:**
Let us consider for $h \in L_k$, $k \in \mathbb{R}_+$ the empirical contrast $\gamma_n(h)$ defined in Equation (3.7). Then we have

$$\widehat{f}_k = \arg\min_{h \in L_k} \gamma_n(h) \quad \text{with} \quad \gamma_n(\widehat{f}_k) = -\|\widehat{f}_k\|^2_{x^{2c-1}},$$

implying that the estimator $\widehat{f}_k$ defined in Equation (3.2) is the MINIMAL CONTRAST ESTIMATOR with respect to the empirical contrast γ_n over the set L_k.

Now for $\chi > 0$ let us define the penalty $\mathrm{pen} : \mathbb{R}_+ \to \mathbb{R}_+$ by $\mathrm{pen}(k) := \chi\sigma_c k n^{-1}$ and an empirical counterpart through $\widehat{\mathrm{pen}}(k) := 2\chi\widehat{\sigma}_c k n^{-1}$ by $\widehat{\sigma}_c := n^{-1}\sum_{j=1}^n X_j^{2(c-1)}$. The definition of pen mimics the behavior of the variance term presented Proposition 3.1.1. Then we can define

$$\widehat{k} := \arg\min_{k \in \mathcal{K}_n} \left(\gamma_n(\widehat{f}_k) + \widehat{\mathrm{pen}}(k)\right) = \arg\min_{k \in \mathcal{K}_n} \left(-\|\widehat{f}_k\|^2_{x^{2c-1}} + \widehat{\mathrm{pen}}(k)\right) \tag{3.8}$$

where $\mathcal{K}_n$, the set of suitable cut-off parameters, is set to $\mathcal{K}_n := [\![n]\!]$. Indeed, a choice of $(k_n)_{n\in\mathbb{N}}$ which leads to a consistent estimator $\widehat{f}_{k_n}$ needs to satisfy $k_n n^{-1} \to 0$ for $n \to \infty$ which on the other hand implies that the set $\mathcal{K}_n$ is rich enough. The following typical result for data-driven method holds true for the choice $\widehat{k}$ defined in Equation (3.8). Here we denote for a measurable function h the essential supremum $\|h\|_\infty := \sup\{k \in \mathbb{R}_+ : \lambda(\{x \in \mathbb{R}_+ : h(x) > k\}) > 0\}$.

Theorem 3.1.13 (Data-driven choice of $k \in \mathbb{R}_+$)**:**
Assume that $f \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$, $\mathbb{E}_f(X_1^{5(c-1)}) < \infty$ and $\|x^{2c-1}f\|_\infty < \infty$. Then for $\chi > 48/\pi$ we get

$$\mathbb{E}^n_f(\|\widehat{f}_{\widehat{k}} - f\|^2_{x^{2c-1}}) \leq 6 \inf_{k \in \mathcal{K}_n} \left(\|f - f_k\|^2_{x^{2c-1}} + \mathrm{pen}(k)\right) + \frac{C_f}{n}$$

where C_f is a positive constant dependent on $\|x^{2c-1}f\|_\infty, \mathbb{E}_f(X_1^{5(c-1)}), \sigma_c$ and χ.

Before proving this result, let us shortly intepret the bound. Theorem 3.1.13 implies that the risk of the proposed fully data-driven estimator $\widehat{f}_{\widehat{k}}$ behaves like the best choice of $k \in \mathcal{K}_n$ which minimizes the sum of the squared bias term and the penalty term, which in turn behaves like the variance term. In other words, due to the richness of the subset $\mathcal{K}_n \subset \mathbb{R}_+$ we can say that the estimator $\widehat{f}_{\widehat{k}}$ behaves similarly to the infimum in (3.6) up to a negligible term. Indeed, after the proof of Theorem 3.1.13 we will derive explicit rates for the data-driven estimator under particular smoothness assumptions, compare Corollary 3.1.18.
The key argument which is used to show Theorem 3.1.13 is the so-called Talagrand inequality, a concentration inequality from the empirical process theory literature. We will present before proving Theorem 3.1.13. The original version is due to Talagrand (1996); the following formulation can be found for example in Klein and Rio (2005).

Lemma 3.1.14 (Talagrand's inequality)**:**
Let $X_1, \dots, X_n$ be independent $\mathcal{Z}$-valued random variables and let

$$\overline{\nu}_h = n^{-1} \sum_{i=1}^{n} [\nu_h(X_i) - \mathbb{E}(\nu_h(X_i))]$$

for ν_h belonging to a countable class $\{\nu_h, h \in \mathcal{H}\}$ of measurable functions. Then,

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}} |\overline{\nu_h}|^2 - 6\Psi^2\right)_+ \leq C\left[\frac{\tau}{n}\exp\left(\frac{-n\Psi^2}{6\tau}\right) + \frac{\psi^2}{n^2}\exp\left(\frac{-Kn\Psi}{\psi}\right)\right] \tag{3.9}$$

with numerical constants $K = (\sqrt{2}-1)/(21\sqrt{2})$ and $C > 0$ and where

$$\sup_{h\in\mathcal{H}}\sup_{z\in\mathcal{Z}} |\nu_h(z)| \leq \psi, \qquad \mathbb{E}(\sup_{h\in\mathcal{H}} |\overline{\nu_h}|) \leq \Psi, \qquad \sup_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \mathbb{V}\mathrm{ar}(\nu_h(X_i)) \leq \tau.$$

Remark 3.1.15:
Keeping the bound (3.9) in mind, let us specify a particular choice for K, i.e. $K \geq \frac{1}{100}$. The next bound is now an immediate consequence,

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}} |\overline{\nu_h}|^2 - 6\Psi^2\right)_+ \leq C\left(\frac{\tau}{n}\exp\left(\frac{-n\Psi^2}{6\tau}\right) + \frac{\psi^2}{n^2}\exp\left(\frac{-n\Psi}{100\psi}\right)\right). \tag{3.10}$$

In the sequel we will make use of the slightly simplified bound (3.10) rather than (3.9).

Proof of Theorem 3.1.13. In general, the proof of for the penalized contrast model selection can be divided in two steps. In the first step, one uses the definition of $\widehat{k}$ in Equation (3.8) and the minimum contrast property stated in Proposition 3.1.12 to show the following Lemma.

Lemma 3.1.16:
Let us define for $h \in B_k := \{h \in L_k : \|h\|^2_{x^{2c-1}} \leq 1\}$ the centered process

$$\overline{\nu}_h := n^{-1}\sum_{j=1}^{n} \nu_h(X_j) - \mathbb{E}_f(\nu_h(X_j)), \text{ where } \nu_h(x) := \frac{1}{2\pi}\int_{\mathbb{R}} x^{c-1+it}\mathcal{M}_c[h](-t)d\lambda(t),$$

for any $x \in \mathbb{R}_+$. Then, under the assumptions of Theorem 3.1.13,

$$\mathbb{E}^n_f(\|\widehat{f}_{\widehat{k}} - f\|^2_{x^{2c-1}}) \leq 6\inf_{k\in\mathcal{K}_n}\left(\|f - f_k\|^2_{x^{2c-1}} + \mathrm{pen}(k)\right) + 8\mathbb{E}^n_f(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}\overline{\nu}_h^2 - p(k))_+)$$
$$+ 2\mathbb{E}^n_f((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+)$$

where $p(k) := \frac{12\sigma_c k}{\pi n}$ and $(a)_+ := \max(a, 0)$ for any $a \in \mathbb{R}$.

The proof of Lemma 3.1.16 can be found in Section 3.1.7.
To finish the proof, it remains to use the Talagrand inequality and basic stochastically calculus, to show that the last two summands in Lemma 3.1.16 are negligible. In order to apply the Talagrand inequality, we first need to split the process since the functions $\nu_h : \mathbb{R}_+ \to \mathbb{C}$ are not bounded on

$\mathbb{R}_+$ contradicting the existence of a constant $\psi \in \mathbb{R}_+$ fulfilling the assumption of Lemma 3.1.14. Therefore, let us decompose the process $\overline{\nu}_h$ into $\overline{\nu}_{h,1}$ and $\overline{\nu}_{h,2}$ for a positive sequence $(d_n)_{n\in\mathbb{N}}$ where

$$\overline{\nu}_{h,1} := n^{-1}\sum_{j=1}^{n} \nu_h(X_j)\mathbb{1}_{(0,d_n)}(X_j^{c-1}) - \mathbb{E}(\nu_h(X_j)\mathbb{1}_{(0,d_n)}(X_j^{c-1}))$$

and $\overline{\nu}_{h,2} := \overline{\nu}_h - \overline{\nu}_{h,1}$. The sequence $(d_n)_{n\in\mathbb{N}}$ will be chosen in Lemma 3.1.17. Then we have

$$\mathbb{E}_f^n\left(\max_{k\in\mathcal{K}_n}\left(\sup_{h\in B_k}\overline{\nu}_h^2 - p(k)\right)_+\right) \leq 2\mathbb{E}_f^n\left(\max_{k\in\mathcal{K}_n}\left(\sup_{h\in B_k}\overline{\nu}_{h,1}^2 - \frac{p(k)}{2}\right)_+\right) + 2\mathbb{E}_f^n\left(\max_{k\in\mathcal{K}_n}\sup_{h\in B_k}\overline{\nu}_{h,2}^2\right).$$

In order to show the Theorem 3.1.13, it remains to control the three upcoming expectations.

Lemma 3.1.17:
Under the assumptions of Theorem 3.1.13 we have

$$(i)\quad \mathbb{E}_f^n\left(\max_{k\in\mathcal{K}_n}\left(\sup_{h\in B_k}\overline{\nu}_{h,1}^2 - \frac{p(k)}{2}\right)_+\right) \leq \frac{C(\|x^{2c-1}f\|_\infty, \sigma_c)}{n},$$

$$(ii)\quad \mathbb{E}_f^n\left(\max_{k\in\mathcal{K}_n}\sup_{h\in B_k}\overline{\nu}_{h,2}^2\right) \leq \frac{C(\mathbb{E}_f(X_1^{5(c-1)}), \sigma_c)}{n} \text{ and}$$

$$(iii)\quad \mathbb{E}_f^n\left((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+\right) \leq \frac{C(\chi, \sigma_c, \mathbb{E}_f(X_1^{4(c-1)}))}{n}.$$

Although the application of the Talagrand inequality follows in general the same steps, we will present it here to highlight the remarkable application of this concentration inequality in the context of data-driven methods.

Proof of Lemma 3.1.17.
Proof of (i): First,

$$\mathbb{E}_f^n\left(\max_{k\in\mathcal{K}_n}\left(\sup_{h\in B_k}\overline{\nu}_{h_1}^2 - \frac{p(k)}{2}\right)_+\right) \leq \sum_{k\in\mathcal{K}_n}\mathbb{E}_f^n\left(\left(\sup_{h\in B_k}\overline{\nu}_{h,1}^2 - \frac{p(k)}{2}\right)_+\right).$$

Now on each summand we will use the Talagrand inequality 3.10 to show the first claim. We want to emphasize that we are able to apply the Talagrand inequality on the sets B_k since B_k has a dense countable subset and due to continuity arguments. Let us start by determining the constant Ψ^2. We define

$$\widetilde{f}_k := \mathcal{M}_c^\dagger[\mathbb{1}_{[-k,k]}n^{-1}\sum_{j=1}^{n} X_j^{c-1+it}\mathbb{1}_{(0,d_n)}(X_j^{c-1})].$$

The function $\widetilde{f}_k$ is, similar to $\widehat{f}_k$, well-defined as the inverse of the Mellin transform of a bounded function with compact support. Additionally, the compact support allows the interchange of integration, leading for any $x \in \mathbb{R}_+$ to

$$\mathbb{E}_f^n(\widetilde{f}_k(x)) = \frac{1}{2\pi}\int_{(-k,k)} x^{-c-it}\mathbb{E}_f(X_1^{c-1+it}\mathbb{1}_{(0,d_n)}(X_1^{c-1}))d\lambda(t)$$

Then, we have for any $h \in B_k$ that $\overline{\nu}_{h,1}^2 = \langle h, \widetilde{f}_k - \mathbb{E}_f^n(\widetilde{f}_k)\rangle_{x^{2c-1}}^2 \leq \|h\|_{x^{2c-1}}^2 \|\widetilde{f}_k - \mathbb{E}_f^n(\widetilde{f}_k)\|_{x^{2c-1}}^2$. Now, since $\|h\|_{x^{2c-1}}^2 \leq 1$ we get

$$\mathbb{E}_f^n\left(\sup_{h\in B_k} \overline{\nu}_{h,1}^2\right) \leq \frac{1}{2\pi n}\int_{(-k,k)} \mathbb{V}\mathrm{ar}_f(X_1^{c-1+it}\mathbb{1}_{(0,d_n)}(X_1^{c-1}))d\lambda(t) \leq \frac{\sigma_c k}{\pi n} =: \Psi^2.$$

Thus $6\Psi^2 = \frac{p(k)}{2}$. Next we consider ψ. Let $x > 0$ and $h \in B_k$. Using the Cauchy-Schwarz inequality we get

$$|\nu_h(x)\mathbb{1}_{(0,d_n)}(x^{c-1})|^2 \leq \frac{d_n^2}{(2\pi)^2}\left(\int_{(-k,k)} |\mathcal{M}_c[h](t)|d\lambda(t)\right)^2 \leq \frac{d_n^2 k}{\pi} =: \psi^2.$$

Next we consider τ. For $h \in B_k$ we can conclude, that

$$\mathbb{V}\mathrm{ar}_f(\nu_h(X_1)\mathbb{1}_{(0,d_n)}(X_1^{c-1})) \leq \mathbb{E}(\nu_h^2(X_1)) \leq \|x^{2c-1}f\|_\infty \|\nu_h\|_{x^{1-2c}}^2.$$

Since $\nu_h(x) = \frac{1}{2\pi}\int_{-k}^k x^{c-1+it}\mathcal{M}_c[h](-t)dt = \mathcal{M}_{1-c}^\dagger[\mathcal{M}_c[h]](x)$ we get by application of the Plancherel identity, Proposition 2.3.5, that

$$\begin{aligned}\mathbb{V}\mathrm{ar}_f(\nu_h(X_1)\mathbb{1}_{(0,d_n)}(X_1^{c-1})) &\leq \|x^{2c-1}f\|_\infty \frac{1}{2\pi}\int_{(-k,k)} |\mathcal{M}_c[h](t)|^2 dt \\ &= \|x^{2c-1}f\|_\infty \|h\|_{x^{2c-1}}^2 \leq \|x^{2c-1}f\|_\infty =: \tau.\end{aligned}$$

Now applying the Talagrand inequality,

$$\begin{aligned}\mathbb{E}_f^n\left(\max_{k\in\mathcal{K}_n}\left(\sup_{h\in B_k}\overline{\nu}_{h,1}^2 - \frac{p(k)}{2}\right)_+\right) &\leq C\sum_{k=1}^n \frac{\|x^{2c-1}f\|_\infty}{n} e^{-\frac{k\sigma_c}{\pi\|x^{2c-1}f\|_\infty}} + \frac{d_n^2 k}{n^2} e^{-\frac{\sqrt{\sigma_c n}}{100 d_n}} \\ &\leq \frac{C(\|x^{2c-1}f\|_\infty, \sigma_c)}{n} + C\sum_{k=1}^n \frac{c_n^2 k}{n^2} e^{-\frac{\sqrt{\sigma_c n}}{100 c_n}}.\end{aligned}$$

By the choice $d_n = \sqrt{\sigma_c n}/(200\log(n))$ we get

$$\sum_{k=1}^n \frac{d_n^2 k}{n^2} e^{-\frac{\sqrt{\sigma_c n}}{100 d_n}} \leq \frac{C(\sigma_c)}{n}\sum_{k=1}^n \frac{k}{n^2} \leq \frac{C(\sigma_c)}{n},$$

which implies the first claim (i).
Now let us show part (ii):
For any $h \in B_k$ we get $\overline{\nu}_{h,2}^2 = \langle h, \widehat{f}_k - \widetilde{f}_k - \mathbb{E}_f^n(\widehat{f}_k - \widetilde{f}_k)\rangle_{x^{2c-1}}^2 \leq \|\widehat{f}_k - \widetilde{f}_k - \mathbb{E}_f^n(\widehat{f}_k - \widetilde{f}_k)\|_{x^{2c-1}}^2$. From this we deduce by application of the Fubini-Tonelli theorem and the Plancherel identity, Proposition 2.3.5, that

$$\begin{aligned}\mathbb{E}_f^n(\max_{k\in\mathcal{K}_n}\sup_{h\in B_k}\overline{\nu}_{h,2}^2) &\leq \frac{1}{2\pi}\mathbb{E}_f^n\left(\int_{-n}^n \left|\mathcal{M}_c[\widehat{f}_n - \widetilde{f}_n](t) - \mathbb{E}_f^n(\mathcal{M}_c[\widehat{f}_n - \widetilde{f}_n](t)\right|^2 d\lambda(t)\right) \\ &= \frac{1}{2n\pi}\int_{-n}^n \mathbb{V}\mathrm{ar}_f(X_1^{c-1+it}\mathbb{1}_{(c_n,\infty)}(X_1^{c-1}))d\lambda(t) \\ &\leq \frac{\mathbb{E}_f(X_1^{2(c-1)}\mathbb{1}_{(c_n,\infty)}(X_1^{c-1}))}{\pi}.\end{aligned}$$

Now for any $p \in \mathbb{R}_+$, $\mathbb{E}_f(X_1^{2(c-1)}\mathbb{1}_{(c_n,\infty)}(X_1^{c-1})) \leq \mathbb{E}_f(X_1^{(2+p)(c-1)}\mathbb{1}_{(c_n,\infty)}(X_1^{c-1}))c_n^{-p}$ implying with $p = 3$ that

$$\mathbb{E}_f^n\left(\max_{k\in\mathcal{K}_n}\sup_{h\in B_k}\overline{\nu}_{h,2}^2\right) \leq c_n^{-3}\mathbb{E}_f(X_1^{(2+3)(c-1)}) \leq \frac{C(\mathbb{E}_f(X_1^{5(c-1)}),\sigma_c)}{n}.$$

To finish the proof of this Lemma, we now show claim (iii):
Let us define the set $\Omega := \{|\widehat{\sigma}_c - \sigma_c| \leq \sigma_c/2\}$. Then on Ω we have $\sigma_c/2 \leq \widehat{\sigma}_c \leq 3/2\sigma_c$ and, since $\widehat{k} \leq n$,

$$\mathbb{E}_f^n((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+) = \frac{\chi}{n}\mathbb{E}_f^n(\widehat{k}(\sigma_c - 2\widehat{\sigma}_c)_+) \leq \chi\mathbb{E}_f^n(\sigma_c - 2\widehat{\sigma}_c)_+\mathbb{1}_{\Omega^c})$$

Now by application of the Cauchy-Schwarz inequality we get

$$\mathbb{E}_f^n(\sigma_c - 2\widehat{\sigma}_c)_+\mathbb{1}_{\Omega^c}) \leq \mathbb{E}_f^n(|\sigma_c - \widehat{\sigma}_c|\mathbb{1}_{\Omega^c}) \leq \mathbb{V}\mathrm{ar}_f^n(\widehat{\sigma}_c)^{1/2}\mathbb{P}_f^n(\Omega_c)^{1/2},$$

where due to the Markov inequality follows $\mathbb{P}_f^n(|\widehat{\sigma}_c - \sigma_c| > \sigma_c/2)^{1/2} \leq 2\mathbb{V}\mathrm{ar}_f^n(\widehat{\sigma}_c)^{1/2}/\sigma_c$. Thus

$$\mathbb{E}_f^n((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+) \leq C(\chi,\sigma_c)\mathbb{V}\mathrm{ar}_f^n(\widehat{\sigma}_c) \leq \frac{C(\chi,\sigma_c,\mathbb{E}_f(X_1^{4(c-1)}))}{n}.$$

□

Combining Lemma 3.1.16 and Lemma 3.1.17 implies the claim of Theorem 3.1.13. □

Now, for densities lying in a Mellin-Sobolev space we can prove the following Corollary.

Corollary 3.1.18:
Let $f \in \mathbb{W}_c^s(L)$, $\|x^{2c-1}f\|_\infty < \infty$ and $\mathbb{E}_f(X_1^{5(c-1)}) < \infty$. Then for any $\chi > 48/\pi$,

$$\mathbb{E}_f^n(\|\widehat{f}_{\widehat{k}} - f\|_{x^{2c-1}}^2) \leq C_f n^{-2s/2s+1}$$

for a positive constant C_f dependent on $L, s, \mathbb{E}_f(X_1^{5(c-1)}), \chi$ and $\|x^{2c-1}f\|_\infty$.

To show that this bound holds true uniformly over a class of function based on the Mellin-Sobolev ellipsoids, we additionally need to control the terms $\mathbb{E}_f(X_1^{5(c-1)})$ and $\|x^{2c-1}f\|_\infty$. While the assumption $\|x^{2c-1}f\|_\infty < \infty$ is rather weak and fulfilled for many densities which have a finite $2(c-1)$ moment, $\sigma_c = \mathbb{E}_f(X_1^{2(c-1)}) < \infty$. On the other hand, the assumption $\mathbb{E}_f(X_1^{5(c-1)})$ could become very crucial if we aim to consider the unweighted $\mathbb{L}^2(\mathbb{R}_+, x^0)$-norm. In this case, we would choose $c = 1/2$ which implies that we need $\mathbb{E}_f(X_1^{-5/2}) < \infty$, implying that the density has to vanish in 0. This might be seen as a boundary condition for the estimation. Nevertheless, the case $c = 1$ implies that both $\sigma_1 = 1 = \mathbb{E}_f(X_1^{5(1-1)}) < \infty$ for all densities. In this particular case, the penalty pen is known and thus does not neet to be estimated.
We close this section by visualizing our estimation strategy.

3.1.5 Numerical results

In the last two subsections we have constructed the estimator $\widehat{f}_k$ for $k \in \mathbb{R}_+$, compare Equation (3.2), we have shown that it is minimax-optimal for an oracle choice k_o, Proposition 3.1.11, and presented a fully-data driven estimator $\widehat{f}_{\widehat{k}}$ in Equation (3.8). Our aim in this subsection is now to visualize the behavior of our estimator $\widehat{f}_{\widehat{k}}$ using a Monte-Carlo simulation. We want to stress out, that a study addressing the behavior of the approximations $(f_k)_{k\in\mathbb{R}_+}$ can be found in Section 2.5. For this simulation study, let us revisit the following examples chosen from Section 2.5.

(i) Beta Distribution: $f(x) = B(2,5)^{-1}x(1-x)^4\mathbb{1}_{(0,1)}(x), x \in \mathbb{R}_+$,

(ii) Log-Gamma Distribution: $f(x) = 5^5\Gamma(5)^{-1}x^{-6}\log(x)^4\mathbb{1}_{(1,\infty)}(x), x \in \mathbb{R}_+$,

(iii) Gamma Distribution: $f(x) = \Gamma(5)^{-1}x^4\exp(-x)\mathbb{1}_{(0,\infty)}(x), x \in \mathbb{R}_+$ and

(iv) Log-Normal Distribution: $f(x) = (0.08\pi x^2)^{-1/2}\exp(-\log(x)^2/0.08)\mathbb{1}_{(0,\infty)}(x), x \in \mathbb{R}_+$.

If $f \in \mathbb{W}^s_c(\mathbb{R}_+)$ for these examples has already been studied in Section 2.5. More precisely, we can calculate the bias for each example $(i)-(iv)$ explicitly and balance the bias and variance term to get the following rates for the upper risk bounds.

	(i)	(ii)	(iii)	(iv)
Rate	$n^{-9/10}$	$n^{-9/10}$	$\log(n)n^{-1}$	$\log(n)^{1/2}n^{-1}$

From a theoretical point of view we would therefore expect that the estimator for the Beta distribution and the Log-Gamma distribution should behave similarly, while for the Gamma distribution, respectively the Log-Normal distribution, the behavior should be noticeable better. It is worth stressing out, that these interpretations are based on an asymptotic point of view. For a fixed sample size $n \in \mathbb{N}$, the multiplicative constant in front of the rate has a much stronger influence on the behavior of the estimator. Doing a preliminary simulation we choose the parameter $\chi = 2$. In Figure 3.1, we presented the choice $c = 1$, since for $c = 1/2$ in the case of (i) we have no theoretical result of the behavior of $\widehat{f}_{\widehat{k}}$ since the assumptions of Theorem 3.1.13 do not hold true. Before considering the behavior of our estimator for changing values of $c \in \mathbb{R}$ we will study the effect of varying sample sizes.
The theory that shows consistency of the estimator $\widehat{f}_{\widehat{k}}$ seem to be visible in the simulation presented in Figure 3.2 as the estimation improves for increasing $n \in \mathbb{N}$. We will now consider the case of varying values of $c \in \mathbb{R}$. In fact, we are considering the cases $c = 0$, $c = 1/2$ and $c = 1$. Therefore, we use examples (ii) and (iv) since both match the assumption of Theorem 3.1.13 for these three choices of $c \in \mathbb{R}$.

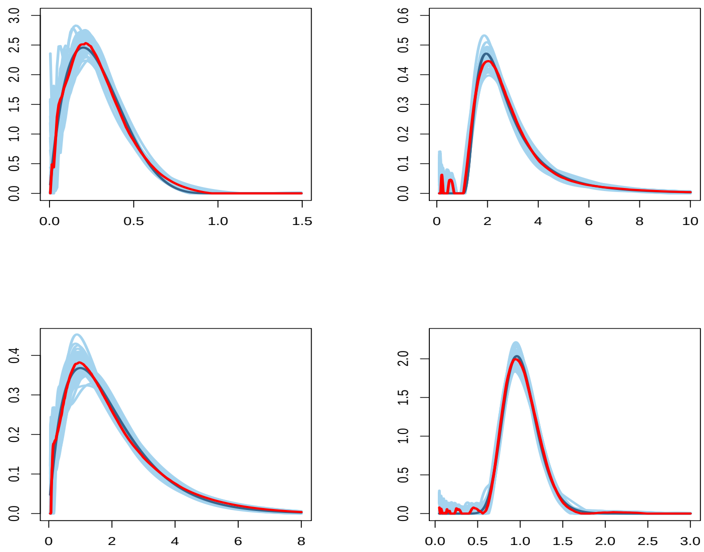

Figure 3.1: Considering the estimator $\widehat{f}_{\widehat{k}}$ and a sample size $n = 500$ the adaptive estimators are depicted for 50 Monte-Carlo simulations in the cases (i) (top, left), (ii) (top, right), (iii) (bottom, left) and (iv) (bottom, right) with $c = 1$. The true density f is given by the dark blue curve while the red curve is the point-wise empirical median of the 50 estimates.

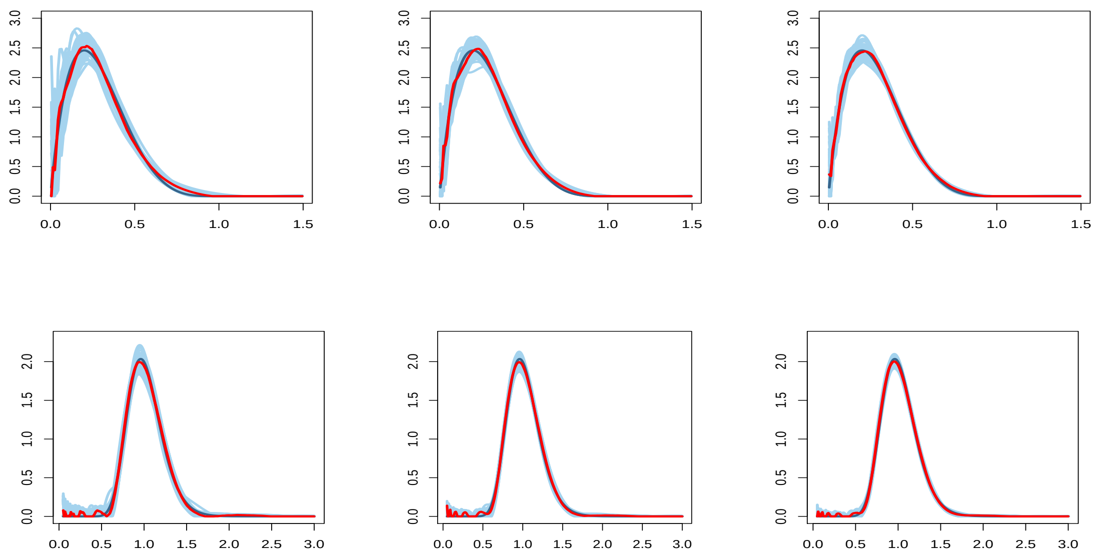

Figure 3.2: Considering the estimator $\widehat{f}_{\widehat{k}}$ with varying sample size $n = 500$ t(left), $n = 1000$ (middle), $n = 2000$ (right) the adaptive estimators are depicted for 50 Monte-Carlo simulations in the cases (i) (top) and (iv) (bottom) with $c = 1$. The true density f is given by the dark blue curve while the red curve is the point-wise empirical median of the 50 estimates.

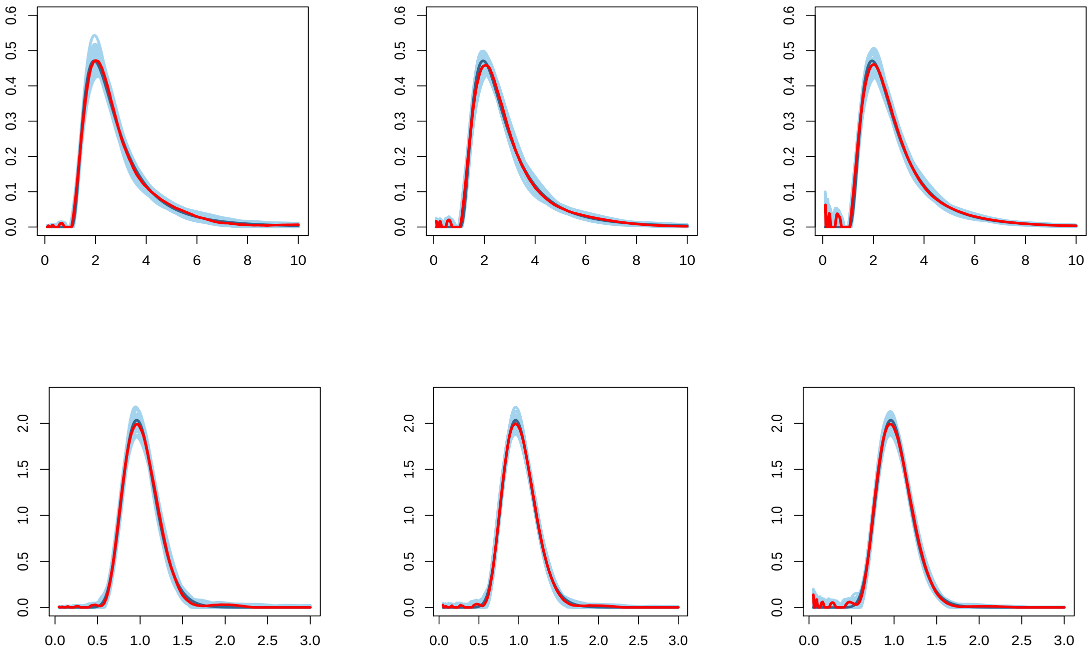

Figure 3.3: Considering the estimator $\widehat{f}_{\widehat{k}}$ and a sample size $n = 1000$ the adaptive estimators are depicted for 50 Monte-Carlo simulations in the cases (ii) (top) and (iv) (bottom) with $c = 0$ (left), $c = 1/2$ (middle) and $c = 1$ (right). The true density f is given by the dark blue curve while the red curve is the point-wise empirical median of the 50 estimates.

Similar to the results in Figure 3.3, we can see that the parameter $c \in \mathbb{R}$ has a strong influence on the approximation behavior of the estimator. For $c = 0$ it seems that the error for values of $x \in \mathbb{R}_+$ close to 0 is rather small, while for bigger values of $x \in \mathbb{R}_+$ the error seems to increase. For the choice $c = 1$ the opposite behavior seems to occur. For $c = 1/2$, none of the both effects can be observed.

3.1.6 Conclusion

In this section we considered the estimation of a density $f : \mathbb{R}_+ \to \mathbb{R}_+$ based on an i.i.d. sample $X_1, \dots, X_n$. We propose an estimator based on the estimation of the Mellin transform of f using the empirical Mellin transform and a regularized inverse Mellin transform. Further, we studied its minimax-optimality over Mellin-Sobolev ellipsoids and proposed a data-driven procedure for the choice of the spectral cut-off parameter. To show that the estimator behaves reasonable in practice, we ended this sections with a Monte Carlo simulation.
Although we are interested in the estimation in the multiplicative measurement error model, we motivated this section by the case of direct observation as a toy example. In the next section, we will use these results and implement them in the context of multiplicative measurement errors.

3.1.7 Proofs

Proof of Lemma 3.1.6. By the fact that $\int_0^1 \psi(x)d\lambda(x) = 0$, we have that any $\delta > 0$ and $\boldsymbol{\theta} \in \{0,1\}^K$ that $\int_0^\infty f_{\boldsymbol{\theta}}(x)d\lambda(x) = 1$.
Due to the construction (3.4) of the function $\psi_{k,K}$, $K \in \mathbb{N}, k \in [\![0, K-1]\!]$, $\psi_{k,K}$ has support on $[1 + k/K, 1 + (k+1)/K]$ which lead to $\psi_{k,K}$ and $\psi_{l,K}$ having disjoint supports if $k \neq l$. For $x \in [1,2]^c$ we have $f_{\boldsymbol{\theta}}(x) = f_o(x) \geq 0$. On the interval $[1,2]$ the function f_o is monotonically decreasing. Furthermore, there is $k_o \in [\![0, K-1]\!]$ such that $x \in [1 + k_o/K, 1 + (k_o+1)/K]$ and hence

$$f_{\boldsymbol{\theta}}(x) = f_o(x) + \theta_{k_o+1}\delta K^{-s}\psi_{k_o,K}(x) \geq f_o(2) - \delta C_{0,\infty}$$

since $\|\psi_{k,K}\|_\infty \leq C_{0,\infty}$ for any $k \in [\![0, K-1]\!]$ and $j \in \mathbb{N}$. Choosing $\delta \leq \delta_o(\psi) = f_o(2)/C_{0,\infty}$ ensures $f_{\boldsymbol{\theta}}(x) \geq 0$ for all $x \in \mathbb{R}_+$. □

Proof of Lemma 3.1.7. Our proof starts with the observation that for all $t \in \mathbb{R}$ we have

$$|\mathcal{M}_c[f_o](t)| = |\exp((c-1+it)^2/2)| = \exp((c-1)^2/2)\exp(-t^2/2),$$

compare Example 2.1.11. Thus for every $s \in \mathbb{N}$ and $c \in \mathbb{R}$ there exists $L_{s,c}$ such that $|f_o|_{s,c}^2 \leq L$ for all $L \geq L_{s,c}$ where $|\,.\,|_{s,c}$ is defined in Definition 3.1.3. Next, we show that there exists a constant $C_{(s,c,\delta)} > 0$ such that $|f_o - f_{\boldsymbol{\theta}}|_{s,c} < C_{(s,c,\delta)}$ which then implies $|f_{\boldsymbol{\theta}}|_{s,c}^2 \leq 2(|f_o - f_{\boldsymbol{\theta}}|_{s,c}^2 + |f_o|_{s,c}^2) \leq 2(C_{(s,c,\delta)} + L_{s,c}) =: L_{s,c,\delta,1}$.
Consider $|f_o - f_{\boldsymbol{\theta}}|_{s,c}$. Let us first define $\Psi_K := \sum_{k=0}^{K-1} \theta_{k+1}\psi_{k,K}$. Then we have $|f_o - f_{\boldsymbol{\theta}}|_{s,c}^2 = \delta^2 K^{-2s}|\Psi_K|_{s,c}^2$ By application of Proposition 2.3.12 and Proposition 2.2.4 we deduce that

$$(1+t^2)^s|\mathcal{M}_c[\Psi_K](t)|^2 = (1+t^2)^s|\mathcal{M}_1[x^{c-1}\Psi_K](t)|^2 = |\mathcal{M}_1[\mathrm{S}^s[x^{c-1}\Psi_k]](t)|^2$$

and thus

$$|\Psi_K|_{s,c}^2 = \int_{-\infty}^{\infty} (1+t^2)^s|\mathcal{M}_c[\Psi_K](t)|^2 d\lambda(t) = (2\pi)\|\mathrm{S}^s[x^{c-1}\Psi_K]\|_{x^{2c-1}}^2$$

by the Parseval identity, see Proposition 2.3.5. By induction, we show that there exist coefficients $(a_j)_{j\in[\![0,s]\!]}$ and $(b_j)_{j\in[\![0,s]\!]}$, possibly dependent on s and c, such that $\mathrm{S}^s[x^{c-1}\Psi_K] = \sum_{j=0}^s a_j x^{b_j}\Psi_K^{(j)}$. Since Ψ_K has compact support in $[1,2]$ we get $|\Psi_K|_{s,c}^2 \leq C_{c,s}\sum_{j=0}^s \|\Psi_K^{(j)}\|_{x^{2c-1}}^2$ and further

$$\begin{aligned}
|\Psi_K|_{s,c}^2 &\leq C_{c,s}\sum_{j=0}^s K^{2j}\sum_{k=0}^{K-1}\int_1^2 x^{2c-1}(\psi^{(j)}(xK-K-k))^2 d\lambda(x) \\
&\leq C_{c,s}\sum_{j=0}^s K^{2j}\sum_{k=0}^{K-1} K^{-1}\|\psi^{(j)}\|_{x^0}^2 \leq C_{c,s}K^{2s}
\end{aligned}$$

where we used that if ψ_k, ψ_l have disjoint support for $l \neq k$, so do their derivatives of any order. Thus, $|f_o - f_{\boldsymbol{\theta}}|_{s,c}^2 \leq C_{(s,c,\delta)}$ and $|f_{\boldsymbol{\theta}}|_s^2 \leq 2(|f_o - f_{\boldsymbol{\theta}}|_s^2 + |f_o|_s^2) \leq 2(C_{(s,c,\delta)} + L_{s,c}) =: L_{s,c,\delta,1}$. Now we still have to show the moment condition. In fact, we can show with the same arguments as before,

$$\int_0^\infty x^{2(c-1)} f_{\boldsymbol{\theta}}(x)dx = e^{2(c-1)^2} + \delta K^{-s}\sum_{k=0}^{K-1}\int_0^\infty x^{2c-2}\psi_{k,K}(x)d\lambda(x) \leq L_{c,\delta,2}.$$

Now choosing $L_{s,c,\delta} := \max(L_{s,c,\delta,1}, L_{c,\delta,2})$ completes the statement.

□

Proof of Lemma 3.1.8.
Using that the functions $(\psi_{k,K})_{k\in[\![0,K-1]\!]}$ with different index k have disjoint supports we get

$$\|f_{\boldsymbol{\theta}} - f_{\boldsymbol{\theta}'}\|^2_{x^{2c-1}} = \delta^2 K^{-2s} \left\| \sum_{k=0}^{K-1} (\theta_{k+1} - \theta'_{k+1})\psi_{k,K} \right\|^2_{x^{2c-1}} = \delta^2 K^{-2s} \rho(\boldsymbol{\theta}, \boldsymbol{\theta}') \|\psi_{0,K}\|^2_{x^{2c-1}}$$

with $\rho(\boldsymbol{\theta}, \boldsymbol{\theta}') := \sum_{j=0}^{K-1} \mathbb{1}_{\{\boldsymbol{\theta}_{j+1} \neq \boldsymbol{\theta}'_{j+1}\}}$ the Hamming distance. The first claim follows by showing that by $\|\psi_{0,K}\|^2_{x^{2c-1}} \geq C_c K^{-1} \|\psi\|^2_{x^0}$ for any $K \in \mathbb{N}$. To do so, we observe that

$$\|\psi_{0,K}\|^2_{x^{2c-1}} = \int_0^\infty x^{2c-1}\psi(xK-K)^2 d\lambda(x) \geq C_c \int_1^2 \psi(xK-K)^2 d\lambda(x) = C_c K^{-1} \|\psi\|^2_{x^0}.$$

Thus $\|f_{\boldsymbol{\theta}} - f_{\boldsymbol{\theta}'}\|^2_{x^{2c-1}} \geq C_c \delta^2 \|\psi\|^2_{x^0} K^{-2s-1} \rho(\boldsymbol{\theta}, \boldsymbol{\theta}')$ for all $K \in \mathbb{N}$.
Now we use the Varshamov-Gilbert Lemma (see Tsybakov (2009)) which states that for $K \geq 8$ there exists a subset $\{\boldsymbol{\theta}^{(0)}, \dots, \boldsymbol{\theta}^{(M)}\}$ of $\{0,1\}^K$ with $\boldsymbol{\theta}^{(0)} = (0, \dots, 0)$ such that $\rho(\boldsymbol{\theta}^{(j)}, \boldsymbol{\theta}^{(k)}) \geq K/8$ for all $j, k \in [\![0, M]\!], j \neq k$ and $M \geq 2^{K/8}$. Applying this leads to $\|f_{\boldsymbol{\theta}^{(j)}} - f_{\boldsymbol{\theta}^{(l)}}\|^2_{x^{2c-1}} \geq C_c \|\psi\|^2_{x^0} \delta^2 K^{-2s}$.
For the second part we have $f_o = f_{\boldsymbol{\theta}^{(0)}}$ and by using

$$\mathrm{KL}(f_{\boldsymbol{\theta}}, f_o) \leq \chi^2(f_{\boldsymbol{\theta}}, f_o) := \int_{\mathbb{R}_+} (f_{\boldsymbol{\theta}}(x) - f_o(x))^2 / f_o(x) d\lambda(x)$$

it is sufficient to bound the χ-squared divergence. We notice that $f_{\boldsymbol{\theta}} - f_o$ has support in $[1,2]$ and f_o is bounded from below on the interval $[1,2]$ by a constant $c_{f_o} > 0$. Now by using the compact support property and a single substitution we get

$$\chi^2(f_{\boldsymbol{\theta}}, f_o) \leq c_{f_o}^{-1} \|f_{\boldsymbol{\theta}} - f_o\|^2 = c_{f_o}^{-1} \delta^2 K^{-2s} \|\Psi_K\|^2_{x^0} = c_{f_o}^{-1} \delta^2 K^{-2s} \sum_{k=0}^{K-1} \frac{\|\psi\|^2_{x^0}}{K} = c_{f_o}^{-1} \delta^2 K^{-2s} \|\psi\|^2_{x^0}.$$

Since $M \geq 2^K$ we have thus $\mathrm{KL}(f_{\boldsymbol{\theta}^{(j)}}, f_{\boldsymbol{\theta}^{(0)}}) \leq C(f_o) \|\psi\|^2_{x^0} \delta^2 \log(M) K^{-2s-1}$. □

Proof of Lemma 3.1.16. First we see that for any $k \in \mathcal{K}_n$ and $h \in L_k$, $\gamma_n(h) = \|h\|^2_{x^{2c-1}} - 2n^{-1} \sum_{j=1}^n \nu_h(X_j)$. From this we can easily see that for $h_1, h_2 \in L_k$,

$$\gamma_n(h_1) - \gamma_n(h_2) = \|h_1 - f\|^2_{x^{2c-1}} - \|h_2 - f\|^2_{x^{2c-1}} - 2\overline{\nu}_{h_1 - h_2}. \tag{3.11}$$

Using Equation (3.11) with $h_1 = \widehat{f}_{\widehat{k}}$ and $h_2 = f_k$ we get

$$\|f - \widehat{f}_{\widehat{k}}\|^2_{x^{2c-1}} \leq \|f - f_k\|^2_{x^{2c-1}} + 2\overline{\nu}_{\widehat{f}_{\widehat{k}} - f_k} + \gamma_n(\widehat{f}_{\widehat{k}}) - \gamma_n(f_k).$$

Since $\widehat{f}_k$ is the minimal contrast estimator, see Proposition 3.1.12, with respect to γ_n and L_k we deduce that $\gamma_n(\widehat{f}_k) \leq \gamma_n(f_k)$ and through the definition of $\widehat{k}$, compare Equation (3.8), we deduce $\gamma_n(\widehat{f}_{\widehat{k}}) + \widehat{\mathrm{pen}}(\widehat{k}) \leq \gamma_n(\widehat{f}_k) + \widehat{\mathrm{pen}}(k)$, which implies that

$$\gamma_n(\widehat{f}_{\widehat{k}}) - \gamma_n(f_k) \leq \gamma_n(\widehat{f}_k) + \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k}) - \gamma_n(f_k) \leq \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k}).$$

And thus

$$\|f - \widehat{f}_{\widehat{k}}\|^2_{x^{2c-1}} \leq \|f - f_k\|^2_{x^{2c-1}} + 2\overline{\nu}_{\widehat{f}_{\widehat{k}} - f_k} + \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k}). \tag{3.12}$$

Let us define $a \vee b := \max(a, b)$ for $a, b \in \mathbb{R}$. Then we have by applying $2ab \leq a^2 + b^2$ that

$$2\overline{\nu}_{\widehat{f}_{\widehat{k}} - f_k} \leq \frac{\|\widehat{f}_{\widehat{k}} - f_k\|^2_{x^{2c-1}}}{4} + 4 \sup_{h \in B_{k \vee \widehat{k}}} \overline{\nu}^2_h \leq \frac{\|\widehat{f}_{\widehat{k}} - f\|^2_{x^{2c-1}} + \|f - f_k\|^2_{x^{2c-1}}}{2} + 4 \sup_{h \in B_{k \vee \widehat{k}}} \overline{\nu}^2_h. \quad (3.13)$$

Combining Equation (3.12) and Equation (3.13) we get

$$\|f - \widehat{f}_{\widehat{k}}\|^2_{x^{2c-1}} \leq 3\|f - f_k\|^2_{x^{2c-1}} + 8 \left(\sup_{h \in B_{\widehat{k} \vee k}} \overline{\nu}^2_h - p(\widehat{k} \vee k) \right)_+ + 8p(\widehat{k} \vee k) + 2\widehat{\mathrm{pen}}(k) - 2\widehat{\mathrm{pen}}(\widehat{k}).$$

Now for $\chi > 48/\pi$ we have $4p(\widehat{k} \vee k) \leq \mathrm{pen}(k) + \mathrm{pen}(\widehat{k})$ and $\mathbb{E}^n_f(\widehat{\mathrm{pen}}(k)) = 2\mathrm{pen}(k)$ which implies

$$\begin{aligned} \mathbb{E}^n_f(\|f - \widehat{f}_{\widehat{f}_k}\|^2_{x^{2c-1}}) \leq & 6 \left(\|f - f_k\|^2_{x^{2c-1}} + \mathrm{pen}(k)\right) + 8\mathbb{E}^n_f \left(\max_{k \in \mathcal{K}_n} \left(\sup_{h \in B_k} \overline{\nu}^2_h - p(k) \right)_+ \right) \\ & + 2\mathbb{E}^n_f((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+). \end{aligned}$$

The claim follows by taking the infimum over $k \in \mathcal{K}_n$ on both sides. □

3.2 Under multiplicative measurement errors

3.2.1 Introduction

We will continue with estimating the unknown density $f : \mathbb{R}_+ \to \mathbb{R}_+$ of a positive random variable X but now given independent and identically distributed (i.i.d.) copies of

$$Y = XU,$$

where X and U are independent of each other and U has a known density $g : \mathbb{R}_+ \to \mathbb{R}_+$. In this setting the density $f_Y : \mathbb{R}_+ \to \mathbb{R}_+$ of Y is given by

$$f_Y(y) = (f * g)(y) = \int_{\mathbb{R}_+} f(x)g(y/x)x^{-1}d\lambda(x), \qquad \text{for } y \in \mathbb{R}_+,$$

where $*$ denotes multiplicative convolution defined in Definition 2.2.2. The estimation of f using an i.i.d. sample $Y_1, \dots, Y_n$ from f_Y is thus an inverse problem called multiplicative deconvolution.

Vardi (1989) and Vardi and Zhang (1992) introduce and study intensively MULTIPLICATIVE CENSORING, which corresponds to the particular multiplicative deconvolution problem with multiplicative error U uniformly distributed on $[0, 1]$. This model is often applied in survival analysis as explained and motivated in van Es et al. (2000). The estimation of the cumulative distribution function of X is considered in Vardi and Zhang (1992) and Asgharian and Wolfson (2005). Series expansion methods are studied in Andersen and Hansen (2001) treating the model as an inverse problem. The density estimation in a multiplicative censoring model is considered in Brunel et al. (2016) using a kernel estimator and a convolution power kernel estimator. Assuming a uniform error distribution on an interval $[1 - \alpha, 1 + \alpha]$ for $\alpha \in (0, 1)$ Comte and Dion (2016) analyze a projection density estimator with respect to the Laguerre basis. Belomestny et al. (2016) study a beta-distributed error U.

In this section, covering all those three variations of the multiplicative censoring model, we consider a density estimator using the Mellin transform and a spectral cut-off regularization of its inverse, which borrows ideas from Belomestny and Goldenshluger (2020). The key to the analysis of the multiplicative deconvolution problem is the convolution theorem of the Mellin transform $\mathcal{M}$, which roughly states $\mathcal{M}[f_Y] = \mathcal{M}[f]\mathcal{M}[g]$ for a density $f_Y = f * g$. Exploiting the convolution theorem Belomestny and Goldenshluger (2020) introduce a kernel density estimator of f allowing more generally X and U to be real-valued. Moreover, they point out that the following widely used naive approach is a special case of their estimation strategy. Transforming the data by applying the logarithm the model $Y = XU$ writes $\log(Y) = \log(X) + \log(U)$. In other words, multiplicative convolution becomes convolution for the $\log$-transformed data. As a consequence, the density of $\log(X)$ is eventually estimated employing usual strategies for non-parametric deconvolution problems, see for example Meister (2009), and then transformed back to an estimator of f. However, it is difficult to interpret regularity conditions on the density of $\log(X)$. Furthermore, the analysis of a global risk of an estimator using this naive approach is challenging as Comte and Dion (2016) pointed out.

Our strategy differs in the following way. Making use of the convolution theorem of the Mellin transform and applying an additional spectral cut-off on the inversion of the Mellin-transform we define a density estimator. We measure the accuracy of the estimator by introducing a global risk in terms of a weighted $\mathbb{L}^2$-norm. Exploiting properties of the Mellin transform we characterize the underlying inverse problem and natural regularity conditions which borrow ideas from the

inverse problems community, Engl et al. (1996). The regularity conditions expressed in the form of Mellin-Sobolev spaces and their relations to the analytical properties of the density f are discussed in more details. The proposed estimator, however, involves a tuning parameter which is selected by a data-driven method. We establish an oracle inequality for the fully-data driven spectral cut-off estimator under fairly mild assumptions on the error density g. Moreover we show that uniformly over Mellin-Sobolev spaces the proposed data-driven estimator is minimax-optimal. Precisely, we state both an upper bound for the mean weighted integrated squared error of the fully-data driven spectral cut-off estimator and a general lower bound for estimating the density f based on i.i.d. copies from $f_Y = f * g$.
The section is organized as follows. In Section 3.2.2 we explain our general estimation strategy by introducing and analyzing our estimator based on the observations $Y_1, \dots, Y_n$. The estimator relies on an inversion of the Mellin transform which we stabilize using a spectral cut-off approach. In Section 3.2.3 we show that the proposed estimator are minimax optimal over Mellin-Sobolev spaces for a large class of error densities g. Based only on the given sample $Y_1, \dots, Y_n$, we will propose in Section 3.2.4 a fully data-driven choice of the upcoming smoothing parameter. Finally, results of a simulation study are reported in section 3.2.5 which visualize the reasonable finite sample performance of our estimator. The proofs of Section 3.2.3 and Section 3.2.4 are postponed to the Appendix.

3.2.2 Estimation strategy

Our aim now is to modify the estimator presented in Equation (3.2) for the case of multiplicative measurement errors, that is we observe an i.i.d. sample $Y_1, \dots, Y_n$ drawn from $f_Y = f * g$. In fact, for a $c \in \mathbb{R}$, with $\mathbb{E}_{f_Y}(Y_1^{c-1}) < \infty$, we can define analogously to Subsection 3.1.2 an unbiased estimator of $\mathcal{M}_c[f_Y]$ using the sample $(Y_j)_{j \in [\![n]\!]}$ by

$$\widehat{\mathcal{M}}_c(t) := n^{-1} \sum_{j=1}^{n} Y_j^{c-1+it}, \quad t \in \mathbb{R}.$$

Applying now the convolution theorem, Proposition 2.2.5, we can state $\mathcal{M}_c[f_Y] = \mathcal{M}_c[f]\mathcal{M}_c[g]$. If we now assume that $\mathcal{M}_c[g](t) \neq 0$ for all $t \in \mathbb{R}$ and that for any $k \in \mathbb{R}_+$ we have $\mathbb{1}_{[-k,k]}/\mathcal{M}_c[g] \in \mathbb{L}^2(\mathbb{R})$, we can again define $\widehat{f}_k := \mathcal{M}_c^\dagger[\mathbb{1}_{[-k,k]}\widehat{\mathcal{M}}_c/\mathcal{M}_c[g]]$. Since $\mathbb{1}_{[-k,k]}\widehat{\mathcal{M}}_c/\mathcal{M}_c[g] \in \mathbb{L}^2(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R})$ we can express $\widehat{f}_k$ explicitly by

$$\widehat{f}_k(x) := \frac{1}{2\pi} \int_{-k}^{k} x^{-c-it} \frac{\widehat{\mathcal{M}}_c(t)}{\mathcal{M}_c[g](t)} d\lambda(t) \quad \text{for any } x \in \mathbb{R}_+. \tag{3.14}$$

To ensure the well-definedness of our estimator $(\widehat{f}_k)_{k \in \mathbb{R}_+}$ we had to impose the moment condition $\mathbb{E}_{f_Y}(Y^{c-1}) < \infty$, which directly forces $\mathbb{E}_g(U^{c-1}) < \infty$, a minor condition on the error density which is captured in the following assumption

$$\forall t \in \mathbb{R} : \mathcal{M}_c[g](t) \neq 0 \text{ and } \forall k \in \mathbb{R}_+ : \frac{\mathbb{1}_{[-k,k]}}{\mathcal{M}_c[g]} \in \mathbb{L}^2(\mathbb{R}). \tag{[G0]}$$

By the construction of the estimator $\widehat{f}_k$ in Equation (3.14), we see that it is an unbiased estimator of the approximations $(f_k)_{k \in \mathbb{R}_+}$ presented in Section 2.5. Similarly to the direct case, we can split the weighted mean integrated squared error into

$$\mathbb{E}^n_{f_Y}(\|\widehat{f}_k - f\|^2_{x^{2c-1}}) = \|f - f_k\|^2_{x^{2c-1}} + \mathbb{E}^n_{f_Y}(\|\widehat{f}_k - f_k\|^2_{x^{2c-1}}).$$

Now for $\mathbb{E}_{f_Y}(Y_1^{2(c-1)}) < \infty$ an application of the Fubini-Tonelli theorem and the Plancherel identity Proposition 2.3.5 allows us to bound the variance term in the following way

$$\mathbb{E}_{f_Y}^n(\|\widehat{f}_k - f_k\|_{x^{2c-1}}^2) = \frac{1}{2\pi}\int_{-k}^{k} \frac{\mathbb{V}\mathrm{ar}_{f_Y}^n(\widehat{\mathcal{M}}_c(t))}{|\mathcal{M}_c[g](t)|^2} d\lambda(t) \leq \frac{\mathbb{E}_{f_Y}(Y_1^{2(c-1)})\Delta_g(k)}{n}$$

where $\Delta_g(k) := (2\pi)^{-1}\int_{-k}^{k}|\mathcal{M}_c[g](t)|^{-2}dt$ for any $k \in \mathbb{R}_+$. This proves directly the multiplicative measurement error model version of Proposition 3.1.1.

Proposition 3.2.1 (Upper bound of the risk):
Let $f \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$, $\sigma_c := \mathbb{E}_{f_Y}(Y_1^{2(c-1)}) < \infty$ and **[G0]** be fulfilled. Then for any $k \in \mathbb{R}_+$

$$\mathbb{E}_{f_Y}^n(\|f - \widehat{f}_k\|_{x^{2c-1}}^2) \leq \|f - f_k\|_{x^{2c-1}}^2 + \frac{\sigma_c}{n}\Delta_g(k)$$

for $\Delta_g(k) := (2\pi)^{-1}\int_{-k}^{k}|\mathcal{M}_c[g](t)|^{-2}dt$ and $\widehat{f}_k$ defined in Equation (3.14).

As we can see, in the case of multiplicative measurement errors, the decay of the Mellin transform of the error density seems to have a strong effect on the variance term. Before we actually have a deeper analysis of the underlying inverse problem using the minimax theory in Section 3.2.3, we want to state that there exists a choice of cut-off parameter, such that our estimator becomes consistent. This is stated in the following Corollary which is a direct consequence of Proposition 3.2.1 and its proof is thus omitted.

Corollary 3.2.2:
Under the assumption of Proposition 3.2.1 choose $(k_n)_{n\in\mathbb{N}}$ such that $k_n \to \infty$ and $\Delta_g(k_n)n^{-1} \to 0$ for $n \to \infty$. Then we have

$$\mathbb{E}_{f_Y}^n(\|f - \widehat{f}_{k_n}\|_{x^{2c-1}}^2) \to 0$$

for $n \to \infty$ and thus $\|f - \widehat{f}_{k_n}\|_{x^{2c-1}}^2 \xrightarrow{\mathbb{P}} 0$.

Remark 3.2.3:
Although the assumption **[G0]** is fulfilled for many error densities the assumption holds not true for the class of so-called OSCILLATING ERROR DENSITIES. In Section 3.4 we consider an estimator based on a ridge approach which can be used for this class of error densities and study its theoretical properties.

3.2.3 Minimax theory

Now we want to develop the minimax theory for the spectral cut-off estimator $\widehat{f}_k$, see (3.14) presented in Subsection 3.2.2. As we have already seen in Proposition 3.2.1 and Proposition 3.4.3, the behavior of the variance term is strongly connected to decay of the Mellin transform of the error density g. In contrast to the direct case presented in Section 3.1, we now have to control both, the bias term and the variance by proposing suitable assumptions on the unknown density f and the error density g. To control the bias term, we will again use the Mellin-Sobolev spaces, respectively the Mellin-Sobolev ellipsoids presented Section 2.5, respectively Subsection 3.1.3. We start by proposing a widely used assumption on the error density.

Noise assumption To determine the growth of the variance term more precisely we will now introduce an additional assumption on the error density. In the context of additive deconvolution problems, compare Fan (1991), densities whose Fourier transform decay polynomially, like in Examples 2.1.6 and 2.1.10, are called SMOOTH ERROR DENSITIES. To stay in this way of speaking we introduce the following definition

Definition 3.2.4 (Smooth error density)**:**
Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\mathbb{E}_g(U^{c-1}) < \infty$ for a $c \in \mathbb{R}$. The we call g a SMOOTH ERROR DENSITY if there exists $c_g, C_g, \gamma \in \mathbb{R}_+$ such that

$$c_g(1+t^2)^{-\gamma/2} \leq |\mathcal{M}_c[g](t)| \leq C_g(1+t^2)^{-\gamma/2} \text{ for all } t \in \mathbb{R}. \qquad \textbf{([G1])}$$

This assumption on the error density was also considered in Belomestny and Goldenshluger (2020). It is clear that **[G1]** implies **[G0]** and that

$$\Delta_g(k) \leq C_g \int_{(-k,k)} (1+|t|^2)^{\gamma} d\lambda(t) \leq C_g k^{2\gamma+1}$$

for a positive constant C_g depending on g.

Remark 3.2.5:
Considering the Examples 2.1.7 and 2.1.4, these error densities do not fulfill assumption **[G1]**. Indeed, these are examples of so-called SUPER SMOOTH error densities which will be considered in Section 3.3.

Minimax bounds Now let us prove upper and lower bounds for the minimax risk of the estimation of the unknown density f based on the i.i.d. sample $Y_1, \dots, Y_n$ over the ellipsoids $\mathbb{D}_c^s(L)$ defined in Section 3.1.3.

Proposition 3.2.6 (Upper bound for the minimax risk)**:**
Let $s, L \in \mathbb{R}_+$ and $c \in \mathbb{R}$. Let additionally $\mathbb{E}_g(U_1^{2(c-1)}) < \infty$ and **[G1]** be fulfilled. Then for the choice $k_o := n^{1/(2s+2\gamma+1)}$

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_{f_Y}^n(\|\widehat{f}_{k_o} - f\|_{x^{2c-1}}^2) \leq C(L, s, g) n^{-2s/(2s+2\gamma+1)}$$

where $C(L, s, g)$ is a positive constant depending on L, s and g.

The proof of Proposition 3.2.6 follows the same argumentation as for Proposition 3.1.4 using that $\Delta_g(k) \leq C_g k^{2\gamma+1}$ for g fulfilling **[G1]**. To show that the presented estimator $\widehat{f}_{k_o}$ are minimax-optimal over the ellipsoids $\mathbb{D}_c^s(L)$, we propose a lower bound for the minimax risk, compare Subsection 3.1.3. For technical reasons, we need the additional assumption that g has a compact support, that is $g(x) = 0$ for any $x > K$ where we say for simplicity that $K = 1$. Further, we need to control the decay of the Mellin transform $\mathcal{M}_c[g]$ of g developed in $\widetilde{c} = 0$ if we consider $c \in (0, 1/2]$ and in $\widetilde{c} = 1/2$ for $c > 1/2$. We collect both conditions in the following assumption

$$\forall x > 1 : g(x) = 0 \text{ and } c_g(1+t^2)^{-\gamma/2} \leq \mathcal{M}_{\widetilde{c}}[g](t)| \leq C_g(1+t^2)^{-\gamma/2}, \quad t \in \mathbb{R}, \qquad \textbf{([G1'])}$$

for constants $c_g, C_g > 0$.

Theorem 3.2.7 (Lower bound for the minimax risk)**:**
Let $s, \gamma \in \mathbb{N}$ and $c \in \mathbb{R}_+$ and assume that **[G1]** and **[G1']** are fulfilled for g with γ. Then there exist constants $C_{g,c}, L_{s,g,c} > 0$ such that for all $L \geq L_{s,g,c}$ and for any estimator $\widehat{f}$ of f based on an i.i.d. sample $Y_1, \dots, Y_n$,

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_{f_Y}^n(\|\widehat{f} - f\|_{x^{2c-1}}^2) \geq C_{c,g} n^{-2s/(2s+2\gamma+1)}.$$

where $C_{c,g}$ is a positive constant depending on c and g.

The proof of Theorem 3.2.7 follows the same steps as the proof of Theorem 3.1.5 with small adaptions to the multiplicative measurement error model. While those adaptions are rather technical and of an elementary nature, we postpone the proof to Section 3.2.7. Let us shortly comment on **[G1']**. If $c > 1/2$ then **[G1']** necessitates $\mathbb{E}_g(U_1^{-1/2}) < \infty$ which is rather weak and fulfilled for many densities. If $c \leq 1/2$ then **[G1']** does not impose an additional assumption since $\mathbb{E}_g(U_1^{-1}) < \infty$ already follows from $\sigma_c = \mathbb{E}_{f_Y}(Y_1^{2(c-1)}) < \infty$. We want to stress out that the family of Beta distribution fulfills both the compact support property and the decay assumption in **[G1]** and **[G1']** with suitable choices of their parameter $(a, b) \in \mathbb{R}_+^2$, compare Example 2.1.6.

Faster rates Again, we can use the same examples as presented in Section 3.1.3 to show that many families of densities possesses a Mellin transform with exponential decay. More precisely, we considered the case of Gamma distribution and Log-Normal distribution where

$$\|f - f_k\|_{x^{2c-1}}^2 \leq C_f \exp(-\alpha k^a)$$

for $\alpha \in \mathbb{R}_+$, dependent on f and $a = 1$ for the Gamma distribution and $a = 2$ for the Log-Normal distribution. Now by choosing $k_n = \log(n)^{1/a}/\alpha^{-1}$ we can ensures that for both cases

$$\inf_{k \in \mathbb{R}_+} \mathbb{E}_{f_Y}^n(\|\widehat{f}_k - f\|_{x^{2c-1}}^2) \leq \mathbb{E}_{f_Y}^n(\|\widehat{f}_{k_n} - f\|_{x^{2c-1}}^2) \leq C_f \frac{\log(n)^{(2\gamma+1)/a}}{n},$$

leading to a faster rate than provided by Proposition 3.1.4 for any choice of $s \in \mathbb{R}_+$, although $f \in \mathbb{W}_c^s(L)$ for any choice of $s \in \mathbb{R}_+$ with L big enough.

3.2.4 Data-driven method

In this section we will present for the spectral cut-off estimator $\widehat{f}_k$ a fully data-driven version and bound its weighted mean integrated squared error.

Let us for $\chi > 0$ define the penalty $\mathrm{pen} : \mathbb{R}_+ \to \mathbb{R}_+$ through $\mathrm{pen}(k) := \chi \sigma_c \Delta_g(k) n^{-1}$ and an empirical counterpart through $\widehat{\mathrm{pen}}(k) := 2\chi\widehat{\sigma}_c \Delta_g(k) n^{-1}$ where $\widehat{\sigma}_c := n^{-1}\sum_{j=1}^n Y_j^{2(c-1)}$. By definition pen mimics the behavior of the variance term presented Proposition 3.2.1. Then we can define

$$\widehat{k} := \arg\min_{k \in \mathcal{K}_n} \left(-\|\widehat{f}_k\|_{x^{2c-1}}^2 + \widehat{\mathrm{pen}}(k) \right) \tag{3.15}$$

where $\mathcal{K}_n$, the set of suitable cut-off parameters, is set $\mathcal{K}_n := [\![K_n]\!]$. Here we choose $K_n \in \mathbb{N}$ as the largest positive integer such that $\Delta_g(K_n) \leq n$, that is $K_n := \max(\{k \in \mathbb{N} : \Delta_g(k) \leq n\})$.

We keep in mind that $(k_n)_{n\in\mathbb{N}}$ needs to suffy $\Delta_g(k_n)n^{-1} \to 0$ for $n \to \infty$ in order for $\widehat{f}_{k_n}$ to be a consistent estimator. Now we can show the following version of Theorem 3.1.13 for the multiplicative measurement error model.

Theorem 3.2.8 (Data-driven choice of $k \in \mathbb{R}_+$)**:**
Assume that $f \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$, $\mathbb{E}_{f_Y}(Y_1^{5(c-1)}) < \infty$, $\|x^{2c-1}g\|_\infty < \infty$ and g is a smooth error density. Then for $\chi > 48$ we get

$$\mathbb{E}^n_{f_Y}(\|\widehat{f}_{\widehat{k}} - f\|^2_{x^{2c-1}}) \leq 6 \inf_{k\in\mathcal{K}_n} \left(\|f - f_k\|^2_{x^{2c-1}} + \mathrm{pen}(k)\right) + \frac{C_{f,g}}{n}$$

where $C_{f,g}$ is a positive constant dependent on $\|x^{2c-1}g\|_\infty, \mathbb{E}_{f_Y}(Y_1^{5(c-1)}), \sigma_c$ and χ.

Proof of Theorem 3.2.8. Again we can divide the proof in two main steps. In the first step, we use the definition of $\widehat{k}$ in Equation (3.15) and the minimum contrast property to show the following Lemma.

Lemma 3.2.9:
Let us define for $h \in B_k := \{h \in L_k : \|h\|^2_{x^{2c-1}} \leq 1\}$ the centered process

$$\overline{\nu}_h := n^{-1}\sum_{j=1}^n \nu_h(Y_j) - \mathbb{E}_{f_Y}(\nu_h(Y_j)), \text{ where } \nu_h(x) := \frac{1}{2\pi}\int_{-\infty}^{\infty} x^{c-1+it}\frac{\mathcal{M}_c[h](-t)}{\mathcal{M}_c[g](t)}d\lambda(t)$$

for any $x \in \mathbb{R}_+$. Then under the assumptions of Theorem 3.2.8,

$$\mathbb{E}^n_{f_Y}(\|\widehat{f}_{\widehat{k}} - f\|^2_{x^{2c-1}}) \leq 6 \inf_{k\in\mathcal{K}_n} \left(\|f - f_k\|^2_{x^{2c-1}} + \mathrm{pen}(k)\right) + 8\mathbb{E}^n_{f_Y}\left(\max_{k\in\mathcal{K}_n}\left(\sup_{h\in B_k}\overline{\nu}_h^2 - p(k)\right)_+\right)$$
$$+ 2\mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+)$$

where $p(k) := \frac{12\sigma_c\Delta_g(k)}{n}$ and $(a)_+ := \max(a, 0)$ for any $a \in \mathbb{R}$.

In the second step, we will use the Talagrand inequality. To do so, we need to decompose the process $\overline{\nu}_h$ into $\overline{\nu}_{h,1}$ and $\overline{\nu}_{h,2}$ where

$$\overline{\nu}_{h,1} := n^{-1}\sum_{j=1}^n \nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{c-1}) - \mathbb{E}_{f_Y}(\nu_h(Y_1)\mathbb{1}_{(0,d_n)}(Y_1^{c-1}))$$

and $\overline{\nu}_{h,2} := \overline{\nu}_h - \overline{\nu}_{h,1}$ where $(d_n)_{n\in\mathbb{N}}$ is a suitable sequence chosen in Lemma 3.2.10. Then we have

$$\mathbb{E}^n_{f_Y}\left(\max_{k\in\mathcal{K}_n}\left(\sup_{h\in B_k}\overline{\nu}_h^2 - p(k)\right)_+\right) \leq 2\mathbb{E}^n_{f_Y}\left(\max_{k\in\mathcal{K}_n}\left(\sup_{h\in B_k}\overline{\nu}_{h,1}^2 - \frac{p(k)}{2}\right)_+\right) + 2\mathbb{E}^n_{f_Y}\left(\max_{k\in\mathcal{K}_n}\sup_{h\in B_k}\overline{\nu}_{h,2}^2\right).$$

In order to show the Theorem it remains to control the three upcoming expectations which is done in the following Lemma.

Lemma 3.2.10:
Under the assumptions of Theorem 3.2.8,

$$(i) \quad \mathbb{E}^n_{f_Y}\left(\max_{k\in\mathcal{K}_n}\left(\sup_{h\in B_k}\overline{\nu}^2_{h,1}-\frac{p(k)}{2}\right)_+\right) \leq) \frac{C(\mathbb{E}_f(X_1^{2(c-1)}), g)}{n};$$

$$(ii) \quad \mathbb{E}^n_{f_Y}\left(\max_{k\in\mathcal{K}_n}\sup_{h\in B_k}\overline{\nu}^2_{h,2}\right) \leq \frac{C(\mathbb{E}_{f_Y}(Y_1^{5(c-1)}), \sigma_c)}{n} \text{ and}$$

$$(iii) \quad \mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k})-\widehat{\mathrm{pen}}(\widehat{k}))_+) \leq \frac{C(\chi, \sigma_c, \mathbb{E}_{f_Y}(Y_1^{4(c-1)}))}{n}$$

Combining Lemma 3.2.9 and Lemma 3.2.10 shows Theorem 3.2.8.

□

Assuming that the density lies in a Mellin-Sobolev ellipsoid, we can deduce directly the following Corollary.

Corollary 3.2.11:
Let $c \in \mathbb{R}$, $s, L \in \mathbb{R}_+$ and $f \in \mathbb{D}^s_c(L)$. Assume further that $\mathbb{E}_{f_Y}(Y_1^{5(c-1)}) < \infty$, $\|gx^{2c-1}\|_\infty < \infty$ and **[G1]** is fulfilled. Then for $\chi > 48$,

$$\mathbb{E}^n_{f_Y}(\|\widehat{f}_{\widehat{k}} - f\|^2_{x^{2c-1}}) \leq C(L, s, g, \mathbb{E}_f(X_1^{5(c-1)}))\, n^{-2s/(2s+2\gamma+1)}$$

where $C(L, s, g, \mathbb{E}_f(X_1^{5(c-1)}))$ is a positive constant depending on L, s, g and $\mathbb{E}_f(X_1^{5(c-1)})$.

3.2.5 Numerical results

We illustrate the behavior of the data-driven spectral cut-off estimator $\widehat{f}_{\widehat{k}}$, presented in equation (3.14) and (3.15). To do so, we use the following examples for the unknown density f

(i) Beta Distribution: $f(x) = B(2,5)^{-1}x(1-x)^4\mathbb{1}_{(0,1)}(x), x \in \mathbb{R}_+$,

(ii) Log-Gamma Distribution: $f(x) = 5^5\Gamma(5)^{-1}x^{-6}\log(x)^4\mathbb{1}_{(1,\infty)}(x), x \in \mathbb{R}_+$,

(iii) Gamma Distribution: $f(x) = \Gamma(5)^{-1}x^4\exp(-x)\mathbb{1}_{(0,\infty)}(x), x \in \mathbb{R}_+$ and

(iv) Log-Normal Distribution: $f(x) = (0.32\pi x^2)^{-1/2}\exp(-\log(x)^2/0.32)\mathbb{1}_{(0,\infty)}(x), x \in \mathbb{R}_+$.

A detailed discussion of these examples in terms of the decay of their Mellin transform and its decay can be found in section 2.5. To visualize the behavior of the estimator, we use the following examples of error densities g

$a)$ Symmetric noise: $g(x) = \mathbb{1}_{(0.5,1.5)}(x), x \in \mathbb{R}_+$,

$b)$ Beta Distribution: $g(x) = 2x\mathbb{1}_{(0,1)}(x), x \in \mathbb{R}_+$,

$c)$ Uniform Distribution: $g(x) = \mathbb{1}_{(0,1)}(x), x \in \mathbb{R}_+$ and

$d)$ Pareto Distribution: $g(x) = \mathbb{1}_{(1,\infty)}(x)x^{-2}, x \in \mathbb{R}_+$.

Here it is worth pointing out that the example $a), c)$ and $d)$ fulfill **[G1]** with $\gamma = 1$ and $b)$ with $\gamma = 2$. By minimizing an integrated weighted squared error over a family of histogram densities with randomly drawn partitions and weights we select $\gamma = 1$ $\chi = 7$ for $\widehat{f}_{\widehat{k}}$. For the case $b)$ we choose $\chi = 5$.

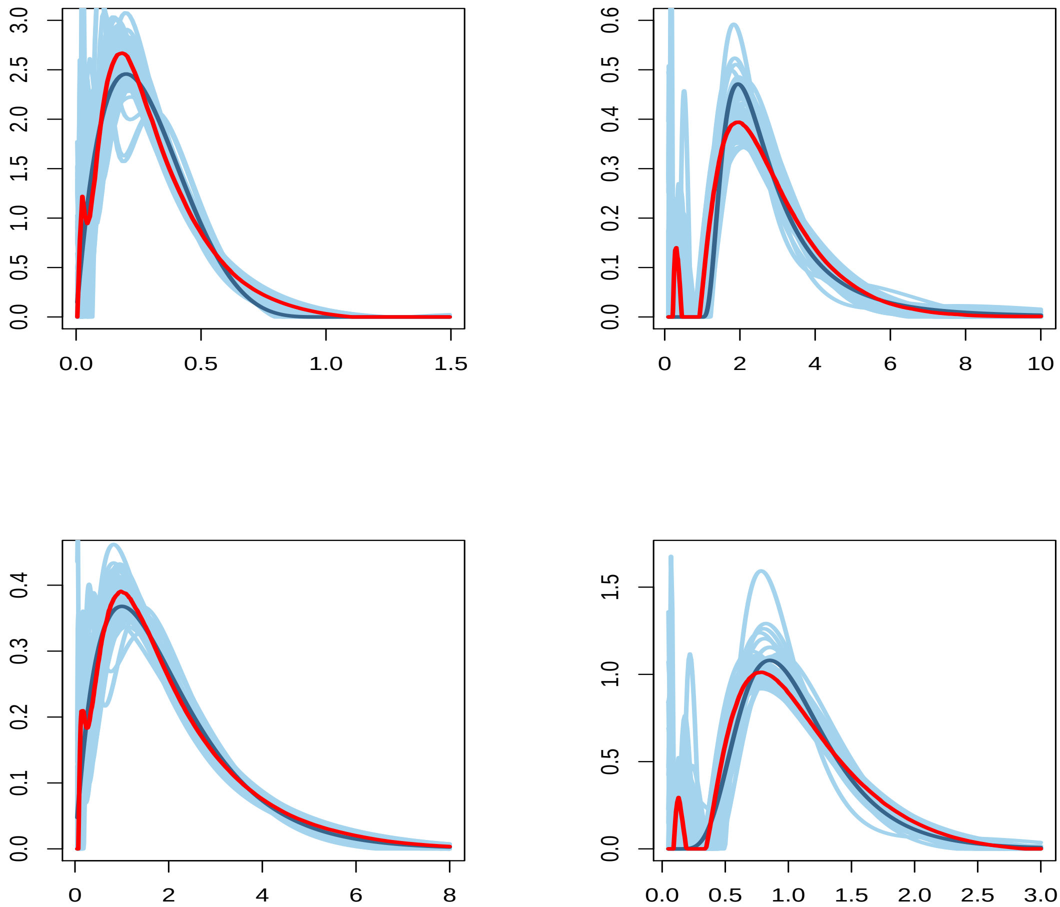

Figure 3.4: Considering the estimator $\widehat{f}_{\widehat{k}}$ and a sample size $n = 500$ the adaptive estimators are depicted for 50 Monte-Carlo simulations in the cases (i) (top, left), (ii) (top, right), (iii) (bottom, left) and (iv) (bottom, right) with $c = 1$ and in the case of a). The true density f is given by the dark blue curve while the red curve is the point-wise empirical median of the 50 estimates.

In Figure 3.4, we presented the choice $c = 1$ since for $c = 1/2$ in the case of (i) we have no theoretical result of the behavior of $\widehat{f}_{\widehat{k}}$ since the assumptions of Theorem 3.2.8 do not hold true. Before considering the behavior of our estimator for changing values of $c \in \mathbb{R}$ we will study the effect of varying sample sizes.

In Figure 3.5, we presented the choice $c = 1$ since for $c = 1/2$ in the case of b) we have no theoretical resutl of the behavior of $\widehat{f}_{\widehat{k}}$ since the assumptions of Theorem 3.2.8 do not hold true. Before considering the behavior of our estimator for changing values of $c \in \mathbb{R}$ we will study the effect of varying sample sizes.

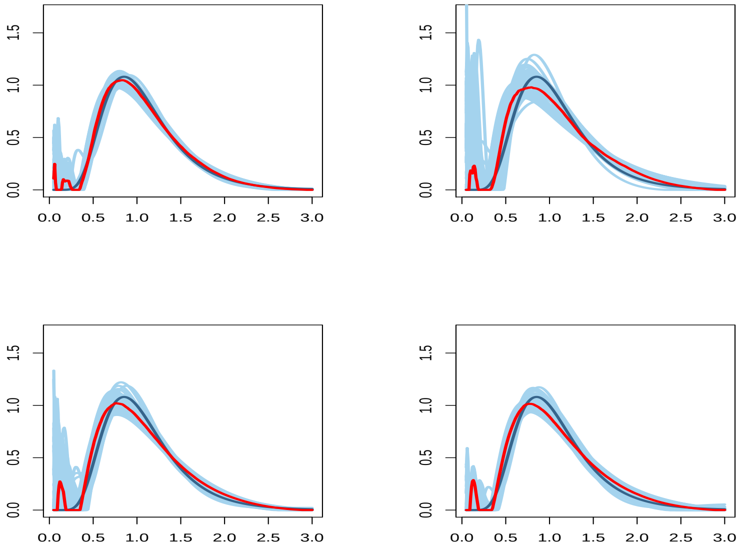

Figure 3.5: Considering the estimator $\widehat{f}_{\widehat{k}}$ and a sample size $n = 500$ the adaptive estimators are depicted for 50 Monte-Carlo simulations in the cases (i) with error a) (top, left), b) (top, right), c) (bottom, left) and d) (bottom, right) and $c = 1$. The true density f is given by the dark blue curve while the red curve is the point-wise empirical median of the 50 estimates.

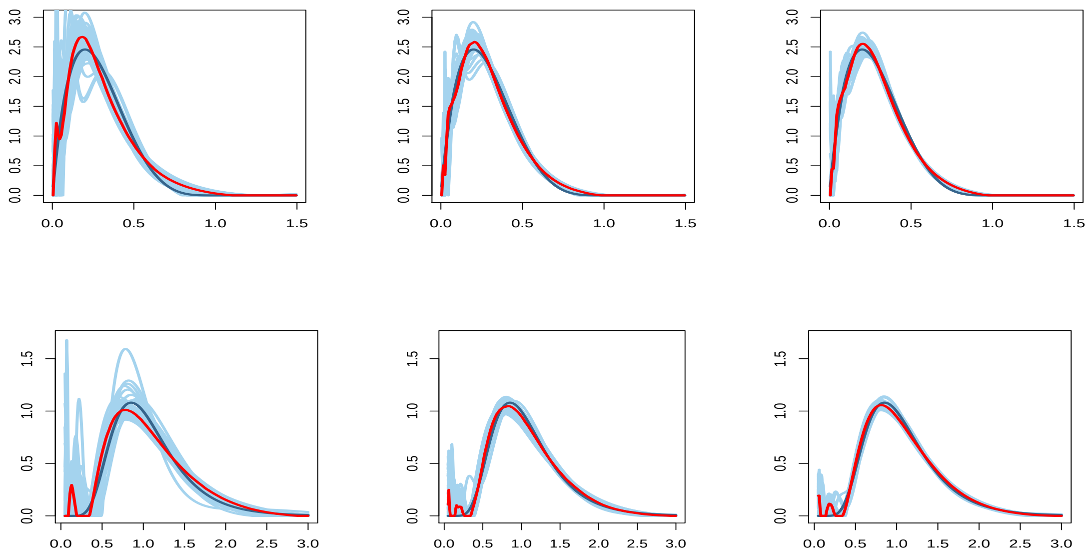

Figure 3.6: Considering the estimator $\widehat{f}_{\widehat{k}}$ with varying sample size $n = 500$ (left), $n = 2000$ (middle), $n = 5000$ (right) the adaptive estimators are depicted for 50 Monte-Carlo simulations in the cases (i) (top) and (iv) (bottom) with $c = 1$ and for a). The true density f is given by the dark blue curve while the red curve is the point-wise empirical median of the 50 estimates.

3.2.6 Conclusion

In this section we have considered the estimation of the density f given an i.i.d. sample of $Y_1, \dots, Y_n$. Throughout the section, we have focused on the case of smooth error densities and showed that our estimation procedure is minimax-optimal and proposed a data-driven choice of the spectral cut-off parameter. We finished the section by illustrating the finite sample properties of our estimator using a Monte-Carlo simulation.
It remains to consider the case of super smooth error densities and error densities which do not fulfill **[G0]**. These two cases are considered in the upcoming sections.

3.2.7 Proofs

Proof of Theorem 3.2.7. Analogously to the proof of Theorem 3.1.5 we construct the family $(f_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\{0,1\}^K}$ through

$$f_{\boldsymbol{\theta}}(x) = f_o(x) + \delta K^{-s-\gamma} \sum_{k=0}^{K-1} \theta_{k+1}\psi_{k,K,\gamma}(x) \tag{3.16}$$

where $f_o(x) = \exp(-x)\mathbb{1}_{\mathbb{R}_+}(x)$ for $c > 1/2$ and $f_o(x) = x\exp(-x)\mathbb{1}_{\mathbb{R}_+}(x)$ for $c \in (0, 1/2]$ and $x \in \mathbb{R}_+$. Here $K \in \mathbb{N}$ is determined later on.
We will now explain how to construct the suitable family of functions $\psi_{k,K,\gamma}$. In fact, let again $\psi \in C_0^\infty(\mathbb{R})$ be a function with support in $[0,1]$ and $\int_0^1 \psi(x)d\lambda(x) = 0$. For any $k \in [\![0, K-1]\!]$ we define the bump-functions $\psi_{k,K}(x) := \psi(xK - K - k)$, $x \in \mathbb{R}$, and define for $j \in \mathbb{N}_0$ the finite constant $C_{j,\infty} := \max(\|\psi^{(l)}\|_\infty, l \in [\![0, j]\!])$. Using then the operator $\mathrm{S} : C_0^\infty(\mathbb{R}) \to C_0^\infty(\mathbb{R})$ with $\mathrm{S}[f](x) = -xf^{(1)}(x)$ for all $x \in \mathbb{R}$ defined in Proposition 2.3.12. Now for $j \in \mathbb{N}$, we define the function $\psi_{k,K,j}(x) := \mathrm{S}^j[\psi_{k,K}](x) = \sum_{i=1}^j c_{i,j}x^i(-K)^i\psi^{(i)}(xK - K - k)$ for $x \in \mathbb{R}_+$ and $c_{i,j} \geq 1$ and let $c_j := \sum_{i=1}^j c_{i,j}$
Until now, we did not give a sufficient condition to ensure that our constructed functions $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \{0,1\}^K\}$ are in fact densities. This condition is given by the following lemma.

Lemma 3.2.12:
Let $0 < \delta < \delta_o(\psi, \gamma) := \exp(-2)2^{-\gamma}(C_{\gamma,\infty}c_\gamma)^{-1}$. Then for all $\boldsymbol{\theta} \in \{0,1\}^K$, $f_{\boldsymbol{\theta}}$ is a density.

Furthermore, we can show that these densities all are contained inside the ellipsoids $\mathbb{D}_c^s(L)$ for L big enough. This is captured in the following lemma.

Lemma 3.2.13:
Let $s \in \mathbb{N}$. Then, there is $L_{s,\gamma,\delta} > 0$ such that f_o and any $f_{\boldsymbol{\theta}}$ as in Equation (3.16) with $\boldsymbol{\theta} \in \{0,1\}^K$, $K \in \mathbb{N}$, belong to $\mathbb{D}_c^s(L_{s,\gamma,\delta})$.

For sake of simplicity we denote for a function $\varphi \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ the multiplicative convolution with g by $\widetilde{\varphi} := (\varphi * g)$. Further we see that for $y \in [0,2]$ holds

$$\widetilde{f}_o(y) = y^{1-2\widetilde{c}} \int_{\mathbb{R}_+} g(x)x^{2\widetilde{c}} \exp(-y/x)d\lambda(x) \geq y^{1-2\widetilde{c}} \int_{\mathbb{R}_+} g(x)x^{-1}\exp(-2/x)d\lambda(x) =: c_{f_o}y^{1-2\widetilde{c}}$$

and thus $\widetilde{f}_o$ is bounded from below by the function $y \mapsto c_{f_o}y^{1-2\widetilde{c}}$ on $[0,2]$. Further we have that $\widetilde{f}_o(2) > 0$ since otherwise $g = 0$ almost everywhere. Exploiting Varshamov-Gilbert's lemma (see Tsybakov (2009)) in Lemma 3.2.14 we show further that there is $M \in \mathbb{N}$ with $M \geq 2^{K/8}$ and a subset $\{\boldsymbol{\theta}^{(0)}, \dots, \boldsymbol{\theta}^{(M)}\}$ of $\{0,1\}^K$ with $\boldsymbol{\theta}^{(0)} = (0, \dots, 0)$ such that for all $j, l \in [\![0, M]\!]$, $j \neq l$ the $\mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$-distance and the Kullback-Leibler divergence are bounded for $K \geq K_o(\gamma, \psi)$.

Lemma 3.2.14:
Let $K \geq K_o(\psi, \gamma, c) \vee 8$. Then there exists a subset $\{\boldsymbol{\theta}^{(0)}, \dots, \boldsymbol{\theta}^{(M)}\}$ of $\{0,1\}^K$ with $\boldsymbol{\theta}^{(0)} = (0, \dots, 0)$ such that $M \geq 2^{K/8}$ and for all $j, l \in [\![0, M]\!]$, with $j \neq l$,

$$(i) \quad \|f_{\boldsymbol{\theta}^{(j)}} - f_{\boldsymbol{\theta}^{(l)}}\|_{x^{2c-1}}^2 \geq \frac{C_c \|\psi^{(\gamma)}\|_{x^0}^2 \delta^2}{16} K^{-2s}$$

$$(ii) \quad \mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}^{(j)}}, \widetilde{f}_{\boldsymbol{\theta}^{(0)}}) \leq \frac{C_{g,c,f_o} \|\psi\|_{x^0}^2}{\widetilde{\log(2)}} \delta^2 \log(M) K^{-2s-2\gamma-1}$$

where KL is the Kullback-Leibler-divergence.

Selecting $K = \lceil n^{1/(2s+2\gamma+1)} \rceil$, it follows

$$\frac{1}{M} \sum_{j=1}^{M} \mathrm{KL}((\widetilde{f}_{\boldsymbol{\theta}^{(j)}})^{\otimes n}, (\widetilde{f}_{\boldsymbol{\theta}^{(0)}})^{\otimes n}) = \frac{n}{M} \sum_{j=1}^{M} \mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}^{(j)}}, \widetilde{f}_{\boldsymbol{\theta}^{(0)}}) \leq C^{(2)}_{\psi,\delta,g,\gamma,f_o,c} \log(M)$$

where $C_{\psi,\delta,g,\gamma,f_o,c} < 1/8$ for all if $\delta \leq \delta_1(\psi, g, \gamma, f_o, c)$ and $M \geq 2$ for $n \geq n_{s,\gamma} := 8^{2s+1} \vee K_o(\gamma, \psi, c)^{2s+2\gamma+1}$. Thereby, we can use Theorem 2.5 of Tsybakov (2009), which in turn for any estimator $\widehat{f}$ of f implies

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{P}^n_{f_Y} \left(\|\widehat{f} - f\|^2_{x^{2c-1}} \geq \frac{C^{(1)}_{\psi,\delta,\gamma,c}}{2} n^{-\frac{2s}{2s+2\gamma+1}} \right) \geq \frac{\sqrt{M}}{1+\sqrt{M}} \left(1 - 1/4 - \sqrt{\tfrac{1}{4\log(M)}}\right)$$
$$\geq 0.07.$$

Note that the constant $C^{(1)}_{\psi,\delta,\gamma,c}$ does only depend on ψ, γ, δ and c, hence it is independent of the parameters s, L and n. The claim of Theorem 3.2.7 follows by using Markov's inequality, which completes the proof. $\square$

Proofs of the lemmata

Proof of Lemma 3.2.12. For any $h \in C_0^\infty(\mathbb{R})$ we can state that $\int_{-\infty}^{\infty} \mathrm{S}[h](x) d\lambda(x) = -[xh(x)]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} h(x) d\lambda(x) = \int_{-\infty}^{\infty} h(x) d\lambda(x)$ and therefore $\int_{-\infty}^{\infty} \mathrm{S}^j[h](x) d\lambda(x) = (-1)^j \int_{-\infty}^{\infty} h(x) d\lambda(x)$ for $j \in \mathbb{N}$. Thus $\int_{-\infty}^{\infty} \psi_{k,K,\gamma}(x) d\lambda(x) = (-1)^\gamma \int_{-\infty}^{\infty} \psi_{k,K}(x) d\lambda(x) = 0$ which implies that for any $\delta > 0$ and $\boldsymbol{\theta} \in \{0,1\}^K$ we have $\int_0^\infty f_{\boldsymbol{\theta}}(x) dx = 1$.
Now due to the construction of Equation (3.16) we easily see that the function $\psi_{k,K}$ has support on $[1 + k/K, 1 + (k+1)/K]$ which lead to $\psi_{k,K}$ and $\psi_{l,K}$ having disjoint supports if $k \neq l$. Here, we want to emphasize that $\mathrm{supp}(\mathrm{S}[h]) \subseteq \mathrm{supp}(h)$ for all $h \in C_0^\infty(\mathbb{R})$. This implies that $\psi_{k,K,\gamma}$ and $\psi_{l,K,\gamma}$ have disjoint supports if $k \neq l$, too. For $x \in [1,2]^c$ we have $f_{\boldsymbol{\theta}}(x) = f_o(x) \geq 0$. Now let us consider the case $x \in [1,2]$. In fact, there is $k_o \in [\![0, K-1]\!]$ such that $x \in [1 + k_o/K, 1 + (k_o+1)/K]$ and hence

$$f_{\boldsymbol{\theta}}(x) = f_o(x) + \theta_{k_o+1} \delta K^{-s-\gamma} \psi_{k_o,K,\gamma}(x) \geq \exp(-2) - \delta 2^\gamma C_{\gamma,\infty} c_\gamma$$

since $\|\psi_{k,K,j}\|_\infty \leq 2^j C_{j,\infty} c_j K^j$ for any $k \in [\![0, K-1]\!]$ and $j \in \mathbb{N}$ where $c_j := \sum_{i=1}^j c_{i,j}$. Now choosing $\delta \leq \delta_o(\psi, \gamma) = \exp(-2) 2^{-\gamma} (C_{\gamma,\infty} c_\gamma)^{-1}$ ensures $f_{\boldsymbol{\theta}}(x) \geq 0$ for all $x \in \mathbb{R}_+$. $\square$

Proof of Lemma 3.2.13. Our proof starts with the observation that for all $t \in \mathbb{R}$ we have $\mathcal{M}_c[f_o](t) = \Gamma(c+it)$ for $c > 1/2$ and $\mathcal{M}_c[f_o](t) = \Gamma(c+1+it)$ for $c \in (0, 1/2]$. Now by applying the Stirling formula, compare Section 2.5, we get thus for every $s \in \mathbb{N}$ there exists $L_{s,c}$ such that $|f_o|_{s,c}^2 \leq L$ for all $L \geq L_{s,c}$.
Next we consider $|f_o - f_{\boldsymbol{\theta}}|_{s,c}$. Let us therefore define first $\Psi_K := \sum_{k=0}^{K-1} \theta_{k+1}\psi_{k,K}$ and $\Psi_{K,j} := \mathrm{S}^j[\Psi_K]$ for an $j \in \mathbb{N}$. Then we have $|f_o - f_{\boldsymbol{\theta}}|_{s,c}^2 = \delta^2 K^{-2s-2\gamma}|\Psi_{K,\gamma}|_{s,c}^2$. By application of Proposition 2.3.12 we deduce that $|\mathcal{M}_c[\Psi_{K,s+\gamma}](t)|^2 = (c+t^2)^s|\mathcal{M}_c[\Psi_{K,\gamma}](t)|^2$ for any $t \in \mathbb{R}$ and thus

$$|\Psi_{K,\gamma}|_{s,c}^2 \leq C_{s,c}\int_{\mathbb{R}} |\mathcal{M}_c[\Psi_{K,s+\gamma}](t)|^2 d\lambda(t) = C_{s,c}\int_{\mathbb{R}_+} x^{2c-1}|\Psi_{K,s+\gamma}(x)|^2 d\lambda(x)$$

by the Parseval identity, compare Proposition 2.3.5. Since $\psi_{k,K}$ have disjoint support for different values of k we follow that $|\Psi_{k,\gamma}|_{s,c}^2 \leq C_{s,c}\sum_{k=0}^{K-1}\theta_{k+1}^2 \int_{\mathbb{R}_+} x^{2c-1}|\mathrm{S}^{\gamma+s}[\psi_{k,K}](x)|^2 d\lambda(x)$. Applying the Jensen inequality and the fact that $\mathrm{supp}(\psi_{k,K}) \subset [1,2]$ leads to

$$\begin{aligned}|\Psi_{k,\gamma}|_s^2 &\leq C_{s,\gamma,c}\sum_{k=0}^{K-1}\sum_{j=1}^{\gamma+s}\int_1^2 x^{2j+1}K^{2j}\psi^{(j)}(xK-K-k)^2 dx \\ &\leq C_{s,\gamma,c}\, K^{2(\gamma+s)}\sum_{k=0}^{K-1}\sum_{j=1}^{\gamma+s} C_{\psi,s,\gamma}^2 K^{-1} \leq C_{s,\gamma,c}K^{2(\gamma+s)}\end{aligned}$$

Thus $|f_o - f_{\boldsymbol{\theta}}|_s^2 \leq C_{s,\gamma,c,\delta}$ and $|f_{\boldsymbol{\theta}}|_s^2 \leq 2(|f_o - f_{\boldsymbol{\theta}}|_s^2 + |f_o|_s^2) \leq 2(C_{s,\gamma,c,\delta} + L_{s,c}) =: L_{s,\gamma,c,\delta}$. □

Proof of Lemma 3.2.14.
Using that the functions $(\psi_{k,K,\gamma})_{k\in[\![0,K-1]\!]}$ with different index k have disjoint supports we get

$$\begin{aligned}\|f_{\boldsymbol{\theta}} - f_{\boldsymbol{\theta}'}\|_{x^{2c-1}}^2 &= \delta^2 K^{-2s-2\gamma}\|\sum_{k=0}^{K-1}(\theta_{k+1} - \theta'_{k+1})\psi_{k,K,\gamma}\|_{x^{2c-1}}^2 \\ &= \delta^2 K^{-2s-2\gamma}\rho(\boldsymbol{\theta},\boldsymbol{\theta}')\|\psi_{0,K,\gamma}\|_{x^{2c-1}}^2\end{aligned}$$

with $\rho(\boldsymbol{\theta},\boldsymbol{\theta}') := \sum_{j=0}^{K-1}\mathbb{1}_{\{\boldsymbol{\theta}_{j+1}\neq\boldsymbol{\theta}'_{j+1}\}}$ the HAMMING DISTANCE. Now the first claim follows by showing that by $\|\psi_{0,K,\gamma}\|_{x^{2c-1}}^2 \geq \frac{C_c K^{2\gamma-1}\|\psi^{(\gamma)}\|_{x^0}^2}{2}$ for K big enough. To do so we observe that

$$\|\psi_{0,K,\gamma}\|_{x^{2c-1}}^2 = \sum_{i,j\in[\![1,\gamma]\!]} c_{j,\gamma}c_{i,\gamma}\int_{\mathbb{R}_+} x^{j+i+2c-1}\psi_{0,K}^{(j)}(x)\psi_{0,K}^{(i)}(x)d\lambda(x).$$

Defining $\Sigma := \|\psi_{0,K,\gamma}\|_{x^{2c-1}}^2 - \int_{\mathbb{R}_+}(x^\gamma\psi_{0,K}^{(\gamma)}(x))^2 x^{2c-1}d\lambda(x)$, we get for $|\Sigma| \leq \frac{C_c K^{2\gamma-1}\|\psi^{(\gamma)}\|_{x^0}^2}{2}$ that

$$\|\psi_{0,K,\gamma}\|_{x^{2c-1}}^2 = \Sigma + \int_{\mathbb{R}_+}(x^\gamma\psi_{0,K}^{(\gamma)}(x))^2 x^{2c-1}d\lambda(x) \geq \Sigma + C_c K^{2\gamma-1}\|\psi^{(\gamma)}\|_{x^0}^2 \geq \frac{C_c K^{2\gamma-1}\|\psi^{(\gamma)}\|_{x^0}^2}{2}. \tag{3.17}$$

It remains to show $|\Sigma| \leq \frac{C_c K^{2\gamma-1}\|\psi^{(\gamma)}\|_{x^0}^2}{2}$. This follows from a comparison of exponents and exploiting the identity $\psi_{0,K}^{(j)}(x) = K^j\psi^{(j)}(xK-K)$, $x \in \mathbb{R}_+$. This is obviously true as soon as $K \geq K_o(\gamma,\psi,c)$, and thus $\|f_{\boldsymbol{\theta}} - f_{\boldsymbol{\theta}'}\|_{x^{2c-1}}^2 \geq \frac{\delta^2 C_c\|\psi^{(\gamma)}\|_{x^0}^2}{2}K^{-2s-1}\rho(\boldsymbol{\theta},\boldsymbol{\theta}')$ for $K \geq K_o(\psi,\gamma,c)$.

Now we use the VARSHAMOV-GILBERT LEMMA (see Tsybakov (2009)) which states that for $K \geq 8$ there exists a subset $\{\boldsymbol{\theta}^{(0)}, \dots, \boldsymbol{\theta}^{(M)}\}$ of $\{0,1\}^K$ with $\boldsymbol{\theta}^{(0)} = (0, \dots, 0)$ such that $\rho(\boldsymbol{\theta}^{(j)}, \boldsymbol{\theta}^{(k)}) \geq K/8$ for all $j, k \in [\![0, M]\!], j \neq k$ and $M \geq 2^{K/8}$. Applying this leads to $\|f_{\boldsymbol{\theta}^{(j)}} - f_{\boldsymbol{\theta}^{(l)}}\|^2_{x^{2c-1}} \geq \frac{C_c \|\psi^{(\gamma)}\|^2_{x^0} \delta^2}{16} K^{-2s}$.
For the second part we have $f_o = f_{\boldsymbol{\theta}^{(0)}}$ and by using $\mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}}, \widetilde{f}_o) \leq \chi^2(\widetilde{f}_{\boldsymbol{\theta}}, \widetilde{f}_o) := \int_{\mathbb{R}_+} (\widetilde{f}_{\boldsymbol{\theta}}(x) - \widetilde{f}_o(x))^2 / \widetilde{f}_o(x) d\lambda(x)$ it is sufficient to bound the χ-squared divergence. We notice that $\widetilde{f}_{\boldsymbol{\theta}} - \widetilde{f}_o$ has support in $[0, 2]$ since $f_{\boldsymbol{\theta}} - f_o$ has support in $[1, 2]$ and g has support in $[0, 1]$. In fact, for $y > 2$ holds $\widetilde{f}_{\boldsymbol{\theta}}(y) - \widetilde{f}_o(y) = \int_y^\infty (f_{\boldsymbol{\theta}} - f_o)(x) x^{-1} g(y/x) d\lambda(x) = 0$. Denote further $\Psi_{K,\gamma} := \sum_{k=0}^{K-1} \theta_{k+1} \psi_{k,K,\gamma} = \mathrm{S}^\gamma[\sum_{k=0}^{K-1} \theta_{k+1} \psi_{k,K}] =: \mathrm{S}^\gamma[\Psi_K]$. Now by using the compact support property and a single substitution we get

$$\chi^2(\widetilde{f}_{\boldsymbol{\theta}}, \widetilde{f}_o) \leq c_{f_o} \|\widetilde{f}_{\boldsymbol{\theta}} - \widetilde{f}_o\|^2_{x^{2\widetilde{c}-1}} = c_{f_o} \delta^2 K^{-2s-2\gamma} \|\widetilde{\Psi}_{K,\gamma}\|^2_{x^{2\widetilde{c}-1}}.$$

Let us now consider $\|\widetilde{\Psi}_{K,\gamma}\|^2_{x^{2\widetilde{c}-1}}$. Now by application of the Parsevaln identity, Proposition 2.3.5, and the convolution theorem, Proposition 2.2.5, that

$$\|\widetilde{\Psi}_{K,\gamma}\|^2_{x^{2\widetilde{c}-1}} = \frac{1}{2\pi} \int_{\mathbb{R}} |\mathcal{M}_{\widetilde{c}}[\widetilde{\Psi}_{K,\gamma}](t)|^2 d\lambda(t) = \frac{1}{2\pi} \int_{\mathbb{R}} |\mathcal{M}_{\widetilde{c}}[g](t)(\widetilde{c} + t^2)^{\gamma/2} \mathcal{M}_{\widetilde{c}}[\Psi_K](t)|^2 d\lambda(t)$$

Then by a using again Plancherel identity and a single change of variable we get

$$\|\widetilde{\Psi}_{K,\gamma}\|^2_{x^{2\widetilde{c}-1}} \leq C_{g,c} \int_{\mathbb{R}} |\mathcal{M}_{\widetilde{c}}[\Psi_K](t)|^2 d\lambda(t) = \frac{C_{g,c}}{2\pi} \|\Psi_K\|^2_{x^{2\widetilde{c}-1}} \leq C_{g,c} \|\Psi_K\|^2_{x^0} = C_{g,c} \|\psi\|^2_{x^0}.$$

Since $M \geq 2^K$ we have thus $\mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}^{(j)}}, \widetilde{f}_{\boldsymbol{\theta}^{(0)}}) \leq \frac{C_{g,c,f_o} \|\psi\|^2}{\log(2)} \delta^2 \log(M) K^{-2s-2\gamma-1}$. □

Proof of Lemma 3.2.9. Defining the empirical contrast for $h \in L_k$,

$$\gamma_n(h) := \|h\|^2_{x^{2c-1}} - 2\frac{1}{2\pi} \int_{\mathbb{R}} \frac{\widehat{\mathcal{M}}_c(t)}{\mathcal{M}_c[g](t)} \mathcal{M}_c[h](-t) d\lambda(t)$$

we see, that $\widehat{f}_k = \underset{h \in L_k}{\arg\min}\ \gamma_n(h)$ with $\gamma_n(\widehat{f}_k) = -\|\widehat{f}_k\|^2_{x^{2c-1}}$. Following the steps of the proof Lemma 3.1.16 we can easily show that

$$\|f - \widehat{f}_{\widehat{k}}\|^2_{x^{2c-1}} \leq 3\|f - f_k\|^2_{x^{2c-1}} + 8 \left(\sup_{h \in B_{\widehat{k} \vee k}} \overline{\nu}_h^2 - p(\widehat{k} \vee k) \right)_+ + 8p(\widehat{k} \vee k) + 2\widehat{\mathrm{pen}}(k) - 2\widehat{\mathrm{pen}}(\widehat{k}).$$

Now for $\chi > 48/\pi$ we have $4p(\widehat{k} \vee k) \leq \mathrm{pen}(k) + \mathrm{pen}(\widehat{k})$ and $\mathbb{E}^n_{f_Y}(\widehat{\mathrm{pen}}(k)) = 2\mathrm{pen}(k)$ which implies

$$\begin{aligned} \mathbb{E}^n_{f_Y}(\|f - \widehat{f}_{\widehat{f}_k}\|^2_{x^{2c-1}}) \leq & 6\left(\|f - f_k\|^2_{x^{2c-1}} + \mathrm{pen}(k)\right) + 8\mathbb{E}^n_{f_Y}\left(\max_{k \in \mathcal{K}_n} \left(\sup_{h \in B_k} \overline{\nu}_h^2 - p(k) \right)_+ \right) \\ & + 2\mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+). \end{aligned}$$

The claim follows then by taking the infimum over $k \in \mathcal{K}_n$ on both sides. □

Proof of Lemma 3.2.10.

Proof of (i): First we see that

$$\mathbb{E}^n_{f_Y}\left(\max_{k\in\mathcal{K}_n}\left(\sup_{h\in B_k}\overline{\nu}^2_{h,1}-\frac{p(k)}{2}\right)_+\right)\leq\sum_{k\in\mathcal{K}_n}\mathbb{E}^n_{f_Y}\left(\left(\sup_{h\in B_k}\overline{\nu}^2_{h,1}-\frac{p(k)}{2}\right)_+\right).$$

Now on each summand we will use the Talagrand inequality 3.10 to show the first claim. We want to emphasize that we are able to apply the Talagrand inequality on the sets B_k since B_k has a dense countable subset and due to continuity arguments. To do so we start to determine the constant Ψ^2. Let us therefore define $\widetilde{f}_k := \mathcal{M}_c^\dagger[\mathbb{1}_{[-k,k]}n^{-1}\sum_{j=1}^n Y_j^{c-1+it}\mathbb{1}_{(0,d_n)}(Y_j^{c-1})/\mathcal{M}_c[g]]$. Then we have for any $h\in B_k$ that $\overline{\nu}^2_{h,1}=\langle h,\widetilde{f}_k-\mathbb{E}^n_{f_Y}(\widetilde{f}_k)\rangle^2_{x^{2c-1}}\leq\|h\|^2_{x^{2c-1}}\|\widetilde{f}_k-\mathbb{E}^n_{f_Y}(\widetilde{f}_k)\|^2_{x^{2c-1}}$. Now since, $\|h\|^2_{x^{2c-1}}\leq 1$ we get

$$\begin{aligned}\mathbb{E}(\sup_{h\in B_k}\overline{\nu}^2_{h,1})\leq\mathbb{E}(\|\widetilde{f}_k-\mathbb{E}(\widetilde{f}_k)\|^2_{x^{2c-1}})&=\frac{1}{2\pi n}\int_{-k}^k\frac{\mathbb{V}\mathrm{ar}(Y_1^{c-1+it}\mathbb{1}_{(0,d_n)}(Y_1^{c-1}))}{|\mathcal{M}_c[g](t)|^2}d\lambda(t)\\&\leq\frac{\sigma_c\Delta_g(k)}{n}=:\Psi^2.\end{aligned}$$

Thus, $6\Psi^2=\frac{p(k)}{2}$. Next we consider ψ. Let $x>0$ and $h\in B_k$. Then using the Cauchy-Schwarz inequality we get

$$|\nu_h(x)\mathbb{1}_{(0,d_n)}(x^{c-1})|^2\leq\frac{d_n^2}{(2\pi)^2}\left(\int_{-k}^k\frac{|\mathcal{M}_c[h](t)|}{|\mathcal{M}_c[g](t)|}d\lambda(t)\right)^2\leq d_n^2\Delta_g(k)=:\psi^2.$$

Next we consider τ. In fact, for $h\in B_k$ we can conclude that

$$\mathbb{V}\mathrm{ar}_f(\nu_h(Y_1)\mathbb{1}_{(0,d_n)}(Y_1^{c-1}))\leq\mathbb{E}_f(\nu_h(Y_1)^2)\leq\|x^{2c-1}f_Y\|_\infty\|\nu_h\|^2_{x^{1-2c}}.$$

Now since

$$\nu_h(x)=\frac{1}{2\pi}\int_{-k}^k x^{c-1+it}\frac{\mathcal{M}_c[h](-t)}{\mathcal{M}_c[g](t)}d\lambda(t)=\mathcal{M}^\dagger_{1-c}\left[\frac{\overline{\mathcal{M}_c[h]}}{\mathcal{M}_c[g]}\right](x),\quad x\in\mathbb{R}_+$$

we get by application of the Plancherel identity, Proposition 2.3.5, that

$$\mathbb{V}\mathrm{ar}_f(\nu_h(Y_1)\mathbb{1}_{(0,d_n)}(Y_1^{c-1}))\leq\|x^{2c-1}f_Y\|_\infty\frac{1}{2\pi}\int_{-k}^k\frac{|\mathcal{M}_c[h](t)|^2}{|\mathcal{M}_c[g](t)|^2}d\lambda(t)\leq\|x^{2c-1}f_Y\|_\infty\|G_k\|_\infty$$

where $G_k(t):=\mathbb{1}_{[-k,k]}(t)|\mathcal{M}_c[g](t)|^{-2}$ for $t\in\mathbb{R}$. Furthermore, we have for any $y\in\mathbb{R}_+$,

$$y^{2c-1}f_Y(y)=\int_{\mathbb{R}_+}\frac{y^{2c-1}}{x^{2c-1}}g(y/x)f(x)x^{2c-2}d\lambda(x)\leq\|x^{2c-1}g\|_\infty\mathbb{E}_f(X_1^{2(c-1)}).$$

We therefore choose $\tau=\|x^{2c-1}g\|_\infty\|G_k\|_\infty\mathbb{E}_f(X_1^{2(c-1)})$. Now applying the Talagrand inequality, we get

$$\begin{aligned}\mathbb{E}^n_{f_Y}\left(\max_{k\in\mathcal{K}_n}\left(\sup_{h\in B_k}\overline{\nu}^2_{h_1}-\frac{p(k)}{2}\right)_+\right)&\leq C(\mathbb{E}_f(X_1^{2(c-1)}),\|x^{2c-1}g\|_\infty)\sum_{k=1}^{K_n}\frac{\|G_k\|_\infty}{n}e^{-\frac{\mathbb{E}_g(U_1^{2(c-1)})\Delta_g(k)}{6\|x^{2c-1}g\|_\infty\|G_k\|_\infty}}\\&\quad+\sum_{k=1}^{K_n}\frac{d_n^2\Delta_g(k)}{n^2}e^{-\frac{\sqrt{\sigma_c n}}{100d_n}}.\end{aligned}$$

Now considering the first summand, we get by application of **[G1]**

$$\sum_{k=1}^{K_n} \frac{\|G_k\|_\infty}{n} e^{-\frac{\mathbb{E}(U_1^{2(c-1)})\Delta_g(k)}{6\|x^{2c-1}g\|_\infty \|G_k\|_\infty}} \le \frac{C_g}{n}\sum_{k=1}^{K_n} k^{2\gamma} e^{-C_g k} \le \frac{C_g'}{n}.$$

By the choice $d_n = \sqrt{\sigma_d n}/(200\log(n))$ and the definition of $\mathcal{K}_n$ we have

$$\sum_{k=1}^{K_n} \frac{d_n^2 \Delta_g(k)}{n^2} e^{-\frac{\sqrt{\sigma_c n}}{100 d_n}} \le \mathbb{E}_f(X_1^{2(c-1)})\frac{C_g}{n}\sum_{k=1}^{K_n}\frac{\Delta_g(k)}{n^2} \le \frac{C(\mathbb{E}_f(X_1^{2(c-1)}), g)}{n}$$

which implies the first claim (i).
Now let us show part (ii):
In fact, for any $h \in B_k$ we get $\overline{\nu}_{h,2}^2 = \langle h, \widehat{f}_k - \widetilde{f}_k - \mathbb{E}^n_{f_Y}(\widehat{f}_k - \widetilde{f}_k)\rangle^2_{x^{2c-1}} \le \|f_k - \widetilde{f}_k - \mathbb{E}^n_{f_Y}(\widehat{f}_k - \widetilde{f}_k)\|^2_{x^{2c-1}}$. From this we deduce by application of the Fubini-Tonelli theorem and the Plancherel identity, Proposition 2.3.5, that

$$\begin{aligned}
\mathbb{E}^n_{f_Y}\left(\max_{k\in\mathcal{K}_n}\sup_{h\in B_k}\overline{\nu}_{h,2}^2\right) &\le \mathbb{E}^n_{f_Y}\left(\frac{1}{2\pi}\int_{-K_n}^{K_n} |\mathcal{M}_c[\widehat{f}_{K_n} - \widetilde{f}_{K_n}](t) - \mathbb{E}^n_{f_Y}(|\mathcal{M}_c[\widehat{f}_{K_n} - \widetilde{f}_{K_n}](t))]^2 d\lambda(t)\right) \\
&= \frac{1}{2n\pi}\int_{-K_n}^{K_n} \frac{\mathbb{V}\mathrm{ar}_{f_Y}(Y_1^{c-1+it}\mathbb{1}_{(d_n,\infty)}(Y_1^{c-1}))}{|\mathcal{M}_c[g](t)|^2} d\lambda(t) \\
&\le \frac{\Delta_g(K_n)}{n}\mathbb{E}_{f_Y}(Y_1^{2(c-1)}\mathbb{1}_{(d_n,\infty)}(Y_1^{c-1})) \le \frac{\mathbb{E}_{f_Y}(Y_1^{(2+3)(c-1)})}{d_n^3}
\end{aligned}$$

since for any $p \in \mathbb{R}_+$ holds $\mathbb{E}_{f_Y}(Y_1^{2(c-1)}\mathbb{1}_{(d_n,\infty)}(Y_1^{c-1})) \le \mathbb{E}_{f_Y}(Y_1^{(2+p)(c-1)})d_n^{-p}$. Thus,

$$\mathbb{E}_{f_Y}\left(\max_{k\in\mathcal{K}_n}\sup_{h\in B_k}\overline{\nu}_{h,2}^2\right) \le d_n^{-3}\mathbb{E}_{f_Y}(Y_1^{(2+3)(c-1)}) \le \frac{C(\mathbb{E}_{f_Y}(Y_1^{5(c-1)}), \sigma_c)}{n}.$$

To finish the proof of this Lemma, we now show claim (iii):
Let us define therefore the set $\Omega := \{|\widehat{\sigma}_c - \sigma_c| \le \sigma_c/2\}$. Then on Ω we have that $\sigma_c/2 \le \widehat{\sigma}_c \le 3/2\sigma_c$ and since $\Delta_g(\widehat{k}) \le n$

$$\mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+) = \frac{\chi}{n}\mathbb{E}^n_{f_Y}(\Delta_g(\widehat{k})(\sigma_c - 2\widehat{\sigma}_c)_+) \le \chi\mathbb{E}(\sigma_c - 2\widehat{\sigma}_c)_+\mathbb{1}_{\Omega_c})$$

Applying the Cauchy-Schwarz inequality we get $\mathbb{E}^n_{f_Y}(\sigma_c - 2\widehat{\sigma}_c)_+\mathbb{1}_{\Omega_c}) \le \mathbb{V}\mathrm{ar}^m_{f_Y}(\widehat{\sigma}_c)^{1/2}\mathbb{P}^n_{f_Y}(\Omega_c)^{1/2}$, where the Markov inequality implies $\mathbb{P}^n_{f_Y}(|\widehat{\sigma}_c - \sigma_c| > \sigma_c/2)^{1/2} \le 2\mathbb{V}\mathrm{ar}^n_{f_Y}(\widehat{\sigma}_c)^{1/2}/\sigma_c$. Thus

$$\mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+) \le C(\chi, \sigma_c)\mathbb{V}\mathrm{ar}^n_{f_Y}(\widehat{\sigma}_c) \le \frac{C(\chi, \sigma_c, \mathbb{E}_{f_Y}(Y_1^{4(c-1)}))}{n}.$$

□

3.3 Under multiplicative measurement errors with super smooth densities

3.3.1 Introduction

Again, we consider the estimation of a density f of a positive random variable X in a multiplicative measurement error model. In other words, we assume to have access to a sample of i.i.d. copies of the random variable

$$Y = XU$$

where U is second positive random variable with density g and is supposed to be stochastically independent of X . This model was intensively motivated in Section 3.2 and studied for the case of smooth error densities.

However, in this section we are interested in the case where the error density g is super smooth. We derive similar results as in the smooth error case and compare these two scenarios. We start by defining the notion of super smooth error densities.

Definition 3.3.1 (Super smooth error density)**:**
Let $g : \mathbb{R}_+ \to \mathbb{R}_+$, such that $\mathbb{E}_g(U^{c-1}) < \infty$ for a $c \in \mathbb{R}$. Then we call g a SUPER SMOOTH ERROR DENSITY, if there exist $c_g, C_g, \lambda, \rho \in \mathbb{R}_+$ and $\gamma \in \mathbb{R}$, such that

$$c_g(1+t^2)^{\gamma/2}\exp(-\lambda|t|^\rho) \le |\mathcal{M}_c[g](t)| \le C_g(1+t^2)^{\gamma/2}\exp(-\lambda|t|^\rho) \text{ for all } t \in \mathbb{R}. \qquad \text{([G2])}$$

The density of the Gamma distribution and the Log-Normal distribution both are super smooth which we capture in the following example and continuation of Example 2.1.7 and Example 2.1.11.

Example 3.3.2 (Super smooth errors: The Gamma distribution)**:**
Let us consider the family of densities $(f_{p,\lambda})_{(p,\lambda)\in\mathbb{R}_+^2}$ of the Gamma distribution where

$$f_{p,\lambda}(x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1}\exp(-\lambda x)\mathbb{1}_{(0,\infty)}(x), \qquad \text{for } p, \lambda, x \in \mathbb{R}_+.$$

Then for $c \in (1-p, \infty)$ and $t \in \mathbb{R}$ we have $|\mathcal{M}_c[f_{p,\lambda}](t)| = \lambda^{1-c}|\Gamma(p+c-1+it)|/\Gamma(p)$, implying with the Stirling formula, compare Belomestny and Goldenshluger (2020), that there exist $c_g, C_g \in \mathbb{R}_+$ such that

$$c_g(1+t^2)^{p+c-3/2}\exp(-|t|\pi/2) \le |\mathcal{M}_c[f_{p,\lambda}](t)| \le C_g(1+t^2)^{p+c-3/2}\exp(-|t|\pi/2).$$

In other words, the error density is super smooth with $\rho = 1, \lambda = \pi/2$ and $\gamma = 2p + 2c - 3$.

Example 3.3.3 (Super smooth errors: The Log-Normal distribution)**:**
Let us consider the family of densities $(f_{\mu,\theta})_{(\mu,\theta)\in\mathbb{R}\times\mathbb{R}_+}$ of the Log-Normal distribution where

$$f_{\mu,\theta}(x) = \frac{1}{\sqrt{2\pi}\theta x}\exp(-(\log(x)-\mu)^2/2\theta^2)\mathbb{1}_{(0,\infty)}(x), \quad \text{for } x \in \mathbb{R}_+.$$

Then for $c \in \mathbb{R}$ and $t \in \mathbb{R}$ we have $|\mathcal{M}_c[f_{\mu,\theta}](t)| = \exp(\mu(c-1) + \frac{\theta^2(c-1)^2}{2})\exp(-\frac{\theta^2 t^2}{2})$ implying that there exist $c_g, C_g \in \mathbb{R}_+$ such that

$$c_g \exp(-\theta^2 t^2/2) \leq |\mathcal{M}_c[f_{p,\theta}](t)| \leq C_g \exp(-\theta^2 t^2/2).$$

In other words, the error density is super smooth with $\rho = 2$, $\lambda = \theta^2/2$ and $\gamma = 0$.

Assuming that the error density g is super smooth, or equivalently spoken assuming that **[G2]** is fulfilled, then **[G0]** is trivially true and we can consider the family of spectral-cut off density estimators $(\widehat{f}_k)_{k\in\mathbb{R}_+}$, defined in (3.14), given by

$$\widehat{f}_k(x) := \frac{1}{2\pi}\int_{[-k,k]} x^{-c-it}\frac{\widehat{\mathcal{M}}_c(t)}{\mathcal{M}_c[g](t)}d\lambda(t), \quad x \in \mathbb{R}_+.$$

For this estimator, we have proven in Proposition 3.2.1 the upper bound in the more general case for **[G0]**. In the next section, we derive an upper bound for the minimax risk of the density estimation with super smooth error densities over Mellin-Sobolev ellipsoids and discuss cases, in which this rate is optimal.

3.3.2 Minimax theory

In the variance term of Proposition 3.2.1, we encounter the function

$$k \mapsto \Delta_g(k) := (2\pi)^{-1}\int_{[-k,k]} \frac{1}{|\mathcal{M}_c[g](t)|^2}d\lambda(t)$$

which characterizes the behavior of the variance term with respect to the cut-off parameter $k \in \mathbb{R}_+$. For the smooth error case, the computation of the integral was rather straight forward. For the super smooth case, the following Lemma helps us to derive an upper bound for Δ_g, whose elementary proof is postponed to the proof section.

Lemma 3.3.4:
Let $\gamma \in \mathbb{R}$ and $\lambda, \rho \in \mathbb{R}_+$. Then there exist constants $C_{\gamma,\lambda,\rho}, c_{\gamma,\lambda,\rho} \in \mathbb{R}_+$ such that for all $y \in \mathbb{R}_+$ holds

$$c_{\gamma,\lambda,\rho}y^{1-2\gamma-\rho}\exp(2\lambda y^\rho) \leq \int_0^y (1+t^2)^{-\gamma}\exp(2\lambda t^\rho)dt \leq C_{\gamma,\lambda,\rho}y^{1-2\gamma-\rho}\exp(2\lambda y^\rho).$$

Applying Lemma 3.3.4 we deduce that under the assumption **[G2]** we have

$$\Delta_g(k) \leq C_g k^{1-2\gamma-\rho}\exp(2\lambda k^\rho) \quad \text{for } k \in \mathbb{R}_+.$$

By similar arguments as in the proof of Proposition 3.2.6 we get the following rate of the density estimation over the Mellin-Sobolev ellipsoids defined in Definition 3.1.3 and below.

Proposition 3.3.5 (Upper bound for the minimax risk):
Let $s, L \in \mathbb{R}_+$ and $c \in \mathbb{R}$. Let additionally $\mathbb{E}_g(U_1^{2(c-1)}) < \infty$ and **[G2]** be fulfilled. Then for the choice $k_o := (\log(n)/4\lambda)^{1/\rho}$

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_{f_Y}^n(\|\widehat{f}_{k_o} - f\|_{x^{2c-1}}^2) \leq C(L, s, g) \log(n)^{-2s/\rho},$$

where $C(L, s, g)$ is a positive constant depending on L, s and g.

It is worth stressing out, that the choice of k_o in Proposition 3.3.5 can be generalized through $k_o = (a \log(n)/(2\lambda))^{1/\rho}$ for any $a \in (0, 1)$. For the sake of simplicity, we have chosen $a = 1/2$. Furthermore, the choice of k_o is independent of the explicit value of the smoothness parameter $s \in \mathbb{R}_+$.
To show that the discovered convergence rate is optimal in the sense, that there exists no estimator based only on $(Y_j)_{j \in [\![n]\!]}$ that achieve a faster converrgence rate than $\log(n)^{-2s/\rho}$, a lower bound result is needed. Here, we stress out that for the case $c = 1$, multiplicative and additive deconvolution estimators can always be expressed through each other using a log-, respectively an exp-, transformation of the data. In Chapter 2 we have seen that the same $\log/\exp$ -transformation of the data describes an isometry between the Sobolev and the Mellin-Sobolev spaces. Thus, the following result for the case $c = 1$ is derived as a direct consequence of its pendant for the additive deconvolution, which can be found in Meister (2009).

Theorem 3.3.6 (Lower bound for the minimax risk):
Let $s \in \mathbb{N}$, assume that g fullfills **[G2]** and

$$\left| \frac{d}{dt} \mathcal{M}_1[g](t) \right| \leq C_g \exp(-\mu |t|^\rho), \quad t \in \mathbb{R} \qquad \text{([G2'])}$$

are fulfilled with $\rho, \mu \in \mathbb{R}_+$. Then there exist constants $C_g, L_{s,g} \in \mathbb{R}_+$, such that for all $L \geq L_{s,g}$ and for any estimator $\widehat{f}$ of f based on an i.i.d. sample $Y_1, \ldots, Y_n$,

$$\sup_{f \in \mathbb{D}_1^s(L)} \mathbb{E}_{f_Y}^n(\|\widehat{f} - f\|_x^2) \geq C_g \log(n)^{-2s/\rho},$$

where $C_{c,g}$ is a positive constant depending on c and g.

Now, Proposition 3.3.5 and Theorem 3.3.6 imply that for the case $c = 1$, the estimator $\widehat{f}_{k_o}$ with $k_o := (\log(n)/4\lambda)^{1/\rho}$ is minimax-optimal for the case of super smooth error densities. It is worth stressing out, that the choice k_o itself is not dependent on the unknown regularity parameter $s \in \mathbb{R}_+$ of the Mellin-Sobolev space $\mathbb{D}_c^s(L)$ and is therefore known, since it only depends on properties of g. Nevertheless, the rate of $\log(n)^{-2s/\rho}$ is rather pessimistic in the sense that for many densities inside the Mellin-Sobolev space, a faster rate can be achieved.

Faster rates Let us revisit the Examples 3.3.2 and 3.3.3. More precisely, we consider the case where g follows a Gamma distribution and f a Log-Normal distribution. Then g is a super smooth error density satisfying

$$\|f - f_k\|_{x^{2c-1}}^2 \leq C_f \exp(-\alpha k^2) \quad \text{and} \quad \mathbb{E}(\|\widehat{f}_k - f_k\|_{x^{2c-1}}^2) \leq C_g \exp(\pi k) n^{-1}$$

for an $\alpha \in \mathbb{R}_+$, depending on f. Now choosing $k_n = \log(n)^{1/2}/\alpha^{-1}$ we have for all $\varepsilon \in (0,1)$ that

$$\inf_{k\in\mathbb{R}_+}\Big(\|f-f_k\|^2_{x^{2c-1}} + \mathbb{E}^n_{f_Y}(\|\widehat{f}_k - f_k\|^2_{x^{2c-1}})\Big) \le \mathbb{E}^n_{f_Y}(\|\widehat{f}_{k_n} - f\|^2_{x^{2c-1}}) \le C_f n^{-1+\varepsilon}.$$

Although both densities f and g lie in $\mathbb{D}^s_c(L)$ for any $s \in \mathbb{R}_+$ and $L \in \mathbb{R}_+$ large enough, a faster rate for an oracle choice of $k_o \in \mathbb{R}_+$ is accessible. In other words, the rate presented in Proposition 3.3.5 is of pessimistic nature. In the following subsection, we present a choice of the parameter $k \in \mathbb{R}_+$ only dependent of the observations.

3.3.3 Data-driven choice

For the sake of simplicity, we only focus on the case of $\rho \ge 1$. For the other cases, similar results to the work of Comte et al. (2006) can be deduced. First we reduce the space of possible cut-off parameters to $\mathcal{K}_n := \{k \in [\![(\log(n)/2\lambda)^{1/\rho}]\!] : \Delta_g(k) \le n\}$. Then we define the model selection method for $\chi > 0$ and $\widehat{\sigma}_Y := \frac{1}{n}\sum_{j\in[\![n]\!]} Y_j^{2(c-1)}$ through

$$\widehat{k} := \arg\min_{k\in\mathcal{K}_n} -\|\widehat{f}_k\|^2_{x^{2c-1}} + \widehat{\mathrm{pen}}(k), \text{ where}\breve{a}\widehat{\mathrm{pen}}(k) := \chi\widehat{\sigma}_Y \frac{k^\rho \Delta_g(k)}{n}. \tag{3.18}$$

Compared to the penalty in (3.15), we see that the term

$$\mathrm{pen}(k) := \mathbb{E}^n_{f_Y}(\widehat{\mathrm{pen}}(k)) = \chi\sigma_Y \frac{k^\rho\Delta_g(k)}{n}$$

is not of the same order as the variance term. This overestimation of the variance term for super-smooth error densities appears commonly in the deconvolution literature, compare Comte et al. (2006). We will first state the bound of the fully data-driven estimator similar to Theorem 3.2.8 and sketch the main steps of the proof. Here, we will point out where precisely the overestimation of the variance is necessary and show that the overestimation does not effect the resulting rate of the fully data-driven estimator.

Theorem 3.3.7 (Data-driven choice of $k \in \mathbb{R}_+$)**:**
Let $f \in \mathbb{L}^2(\mathbb{R}^2_+, x^{2c-1})$, $\mathbb{E}_{f_Y}(Y_1^{5(c-1)}) < \infty$ and **[G2]** holds true. Then there exists $\chi_0 \in \mathbb{R}_+$ such that for all $\chi \ge \chi_0$

$$\mathbb{E}^n_{f_Y}(\|\widehat{f}_{\widehat{k}} - f\|^2_{x^{2c-1}}) \le 3\inf_{k\in\mathcal{K}_n}(\|f-f_k\|^2_{x^{2c-1}} + \mathrm{pen}(k)) + \frac{C(\mathbb{E}_{f_Y}(Y_1^{5(c-1)}, \chi, g)}{n},$$

where $C(g,\chi,\mathbb{E}_{f_Y}(Y_1^{5(c-1)})) > 0$ is a positive constant depending on g, χ and $\mathbb{E}_{f_Y}(Y_1^{5(c-1)})$.

Proof of Theorem 3.3.7. The proof of the theorem can be reduced in the following three Lemmas.

Lemma 3.3.8:
Under the assumptions of Theorem 3.3.7 holds

$$\mathbb{E}^n_{f_Y}(\|\widehat{f}_{\widehat{k}} - f\|^2_{x^{2c-1}}) \le 3(\|f-f_k\|^2_{x^{2c-1}} + \mathrm{pen}(k)) + 11\mathbb{E}^n_{f_Y}\Big(\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}} - \frac{1}{12}\mathrm{pen}(\widehat{k})\Big)_+ + \mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - 2\widehat{\mathrm{pen}}(\widehat{k}))_+)$$

The theorem follows by applying the following two lemmata and taking the infimum over $k \in \mathcal{K}_n$.

Lemma 3.3.9:
Under the assumptions of Theorem 3.3.7 we get

$$\mathbb{E}^n_{f_Y}\left(\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c1}} - \frac{1}{12}\mathrm{pen}(\widehat{k})\right)_+ \leq \frac{C(g, \mathbb{E}_{f_Y}(Y_1^{4(c-1)})}{n}.$$

where $C(g) > 0$ is a positive constant only depending on g.

Lemma 3.3.10:
Under the assumptions of Theorem 3.3.7 we get

$$\mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - 2\widehat{\mathrm{pen}}(\widehat{k}))_+) \leq \frac{C(\chi, g, \mathbb{E}_{f_Y}(Y_1^{5(c-1)})}{n},$$

where $C(\chi, g, \mathbb{E}_{f_Y}(Y_1^{5(c-1)})) \in \mathbb{R}_+$ is a positive constant only depending on χ, g, σ_Y and $\mathbb{E}_{f_Y}(Y_1^{4(c-1)})$.

□

We will now directly prove Lemma 3.3.9 to observe where the overestimation of the variance term is necessary compared to the smooth error case. The proofs of Lemma 3.3.8 and Lemma 3.3.10 are postponed to the appendix.

Proof of Lemma 3.3.9. First we see that

$$\mathbb{E}^n_{f_Y}\left(\left(\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}} - \frac{1}{12}\mathrm{pen}(\widehat{k})\right)_+\right) \leq \mathbb{E}^n_{f_Y}\left(\max_{k\in\mathcal{K}_n}\left(\|\widehat{f}_k - f_k\|^2_{x^{2c-1}} - \frac{1}{12}\mathrm{pen}(k)\right)_+\right).$$

Define $B_k := \{h \in S_k : \|h\|_{x^{2c-1}} = 1\}$. Furthermore we set

$$\|\widehat{f}_k - f_k\|^2_{x^{2c-1}} = \sup_{h\in B_k}\langle \widehat{f}_k - f_k, h\rangle^2_{x^{2c-1}} =: \sup_{h\in B_k}\overline{\nu}_h^2,$$

where

$$\bar{\nu}_h := n^{-1}\sum_{j\in[\![n]\!]}(\nu_h(Y_j) - \mathbb{E}_{f_Y}(\nu_h(Y_j))) \text{ and } \nu_h(Y_j) := \frac{1}{2\pi}\int_{[-k,k]}\frac{Y_j^{c-1+it}}{\mathcal{M}_c[g](t)}\mathcal{M}_c[h](t)d\lambda(t).$$

We are aiming to apply the Talagrand inequality, namely Lemma 3.1.14. We therefore split, for a sequence $(d_n)_{n\in\mathbb{N}}$ specified afterwards, the process again in the following way

$$\begin{aligned}\bar{\nu}_{h,1} :=& n^{-1}\sum_{j\in[\![n]\!]}\nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{c-1}) - \mathbb{E}_{f_Y}(\nu_h(Y_1)\mathbb{1}_{(0,d_n)}(Y_1^{c-1}))\\ \text{and } \bar{\nu}_{h,2} :=& n^{-1}\sum_{j\in[\![n]\!]}\nu_h(Y_j)\mathbb{1}_{(d_n,\infty)}(Y_j^{c-1}) - \mathbb{E}_{f_Y}(\nu_h(Y_1)\mathbb{1}_{(d_n,\infty)}(Y_1^{c-1}))\end{aligned}$$

to get

$$\mathbb{E}^n_{f_Y}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}|\bar{\nu}_h|^2-\frac{1}{12}\mathrm{pen}(k))_+)\leq 2\mathbb{E}^n_{f_Y}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}|\bar{\nu}_{h,1}|^2-\frac{1}{24}\mathrm{pen}(k))_++|\bar{\nu}_{h,2}|^2)$$
$$:=M_1+M_2.$$

We will now consider the two summands M_1 and M_2 separately.
To bound the M_1 we will use the Talagrand inequality 3.1.14 on the term $\mathbb{E}^n_{f_Y}(\sup_{t\in B_k}|\bar{\nu}_{h,1}|^2-\frac{1}{24}\widehat{\mathrm{pen}}(k))_+$. Indeed, we have

$$M_1\leq\sum_{k\leq K_n}\mathbb{E}^n_{f_Y}(\sup_{t\in B_k}|\bar{\nu}_{h,1}|^2-\frac{1}{24}\mathrm{pen}(k))_+,$$

which will be used to show the claim. We want to emphasize that we are able to apply the Talagrand inequality on the sets B_k since B_k has a dense countable subset and due to continuity arguments. In order to apply Talagrand's inequality, we need to find the constants Ψ,ψ and τ such that

$$\sup_{h\in B_k}\sup_{y\in\mathbb{R}_+}|\nu_h(y)\mathbb{1}_{(0,d_n)}(y^{c-1})|\leq\psi;\quad \mathbb{E}_{f_Y^n}(\sup_{h\in B_k}|\bar{\nu}_{h,1}|)\leq\Psi;$$
$$\sup_{h\in B_k}\frac{1}{n}\sum_{j\in[\![n]\!]}\mathbb{V}\mathrm{ar}_{f_Y}(\nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{c-1}))\leq\tau.$$

We start to determine the constant Ψ^2. Let us define $\widetilde{\mathcal{M}}_c(t):=n^{-1}\sum_{j\in[\![n]\!]}Y_j^{c-1+it}\mathbb{1}_{(0,d_n)}(Y_j^{c-1})$ as an unbiased estimator of $\mathcal{M}_c[f_Y\mathbb{1}_{(0,d_n)}(y^{c-1})](t)$ and

$$\widetilde{f}_k(x):=\frac{1}{2\pi}\int_{[-k,k]}x^{-c-it}\frac{\widetilde{\mathcal{M}}(t)}{\mathcal{M}_c[g](t)}d\lambda(t),$$

where $n^{-1}\sum_{j\in[\![n]\!]}\nu_h(Y_j)\mathbb{1}_{(0,c_n)}(Y_j^{c-1})=\langle\widetilde{f}_k,h\rangle_{x^{2c-1}}$. Thus, we have for any $h\in B_k$ that $\bar{\nu}_{h,1}^2=\langle h,\widetilde{f}_k-\mathbb{E}^n_{f_Y}(\widetilde{f}_k)\rangle^2_{x^{2c-1}}\leq\|h\|^2_{x^{2c-1}}\|\widetilde{f}_k-\mathbb{E}^n_{f_Y}(\widetilde{f}_k)\|^2_{x^{2c-1}}$. Since $\|h\|_{x^{2c-1}}\leq 1$, we get

$$\mathbb{E}^n_{f_Y}(\sup_{h\in B_k}\bar{\nu}_{h,1}^2)\leq\mathbb{E}^n_{f_Y}(\|\widetilde{f}_k-\mathbb{E}^n_{f_Y}(\widetilde{f}_k)\|^2)=\frac{1}{2\pi}\int_{[-k,k]}\frac{\mathbb{V}\mathrm{ar}^n_{f_Y}(\widetilde{\mathcal{M}}(t))}{|\mathcal{M}_c[g](t)|^2}d\lambda(t)\leq\sigma_Y\frac{\Delta_g(k)}{n},$$

as $\mathbb{V}\mathrm{ar}^n_{f_Y}(\widetilde{\mathcal{M}}(t))\leq n^{-1}\mathbb{E}_{f_Y}(Y_1^{2(c-1)}\mathbb{1}_{(0,d_n)}(Y_1^{c-1})\leq\sigma_Y n^{-1}$. We obtain

$$\mathbb{E}^n_{f_Y}(\sup_{h\in B_k}\bar{\nu}_{h,1}^2)\leq\sigma_Y\frac{\Delta_g(k)}{n}=:\Psi^2.$$

Next we consider ψ. Let $y\in\mathbb{R}_+$ and $h\in B_k$. Then using the Cauchy-Schwarz inequality,

$$|\nu_h(y)\mathbb{1}_{(0,d_n)}(y^{c-1})|^2=(2\pi)^{-2}d_n^2|\int_{[-k,k]}y^{it}\frac{\mathcal{M}_c[h](-t)}{\mathcal{M}_c[g](t)}d\lambda(t)|^2$$
$$\leq(2\pi)^{-2}d_n^2\int_{[-k,k]}|\mathcal{M}_c[g](t)|^{-2}d\lambda(t)\leq d_n^2\Delta_g(k)=:\psi^2$$

since $|y^{it}|=1$ for all $t\in\mathbb{R}$. For τ we use the crude bound

$$\sup_{h\in B_k}\frac{1}{n}\sum_{j=1}^n\mathbb{V}\mathrm{ar}_{f_Y}(\nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{c-1})\leq\frac{1}{2\pi}\int_{[-k,k]}\frac{\mathbb{V}\mathrm{ar}_{f_Y}(Y_1^{c-1+it}\mathbb{1}_{(0,d_n)}(Y_j^{c-1})}{|\mathcal{M}_c[g](t)|^2}d\lambda(t)\leq n\Psi^2,$$

and set $\tau=n\Psi^2$.

Remark 3.3.11 (Overestimation of the variance term):
In the proof of Lemma 3.2.10, we have used the same choice of Ψ^2 while we have chosen

$$\tau = \|x^{2c-1}g\|_\infty \|G_k\|_\infty \mathbb{E}_f(X_1^{2(c-1)}), \quad G_k := \mathbb{1}_{(-k,k)}|\mathcal{M}_c[g]|^{-2}.$$

In the smooth error case **[G1]** we have then deduced that

$$c_g k \leq \frac{n\Psi^2}{\tau} \leq C_g k,$$

implying that the sequence $(C_g k^{2\gamma}\exp(-C_g k))_{k\in\mathbb{N}}$ is summable. For the super smooth case **[G2]** for $\gamma < 0$ we have for any $k \in \mathbb{R}_+$

$$\|G_k\|_\infty \leq \sup_{t\in(0,k)} C_g(1+t^2)^{-\gamma}\exp(2\lambda t^\rho) \leq C_g k^{-2\gamma}\exp(2\lambda k^\rho),$$

respectively $\|G_k\|_\infty \geq c_g k^{-2\gamma}\exp(2\lambda k^\rho)$. Now Lemma 3.3.4 leads to

$$\frac{n\Psi^2}{\tau} \leq C_g \frac{k^{1-\rho-2\gamma}\exp(2\lambda k^\rho)}{k^{-2\gamma}\exp(2\lambda k^\rho)} = C_g k^{1-\rho},$$

which is decreasing in $k \in \mathbb{R}_+$ implying that the first summand in Lemma 3.1.14 might not be bounded w.r.t. $k \in \mathbb{R}_+$ in the case of super smooth error densities. A more general version of the Talagrand inequality is thus needed which on the other hand implies the need of the overestimation of the variance of the estimator.

The following inequality is due to Talagrand (1996), the formulation of the first part can be found for example in Klein and Rio (2005).

Lemma 3.3.12 (Talagrand's inequality):
Let $Z_1, \dots, Z_n$ be independent $\mathcal{Z}$-valued random variables and let $\bar{\nu}_h = n^{-1}\sum_{i=1}^n [\nu_h(Z_i) - \mathbb{E}(\nu_h(Z_i))]$ for ν_h belonging to a countable class $\{\nu_h, h \in \mathcal{H}\}$ of measurable functions. Then, for all $\varepsilon \in \mathbb{R}_+$

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}} |\bar{\nu}_h|^2 - 2(1+2\varepsilon)\Psi^2\right)_+ \leq C\left[\frac{\tau}{n}\exp\left(\frac{-K_1\varepsilon n\Psi^2}{\tau}\right) + \frac{\psi^2}{C_\varepsilon^2 n^2}\exp\left(\frac{-C_\varepsilon K_2\sqrt{\varepsilon}n\Psi}{\psi}\right)\right] \tag{3.19}$$

with numerical constants $C_\varepsilon := \sqrt{1+\varepsilon} - 1$ and $C \in \mathbb{R}_+$ and where

$$\sup_{h\in\mathcal{H}}\sup_{z\in\mathcal{Z}} |\nu_h(z)| \leq \psi, \qquad \mathbb{E}(\sup_{h\in\mathcal{H}} |\bar{\nu}_h|) \leq \Psi, \qquad \sup_{h\in\mathcal{H}} \frac{1}{n}\sum_{j=1}^n \mathbb{V}\mathrm{ar}(\nu_h(Z_i)) \leq \tau.$$

It is clear to see that the Lemma 3.1.14 follows from Lemma 3.3.12 for $\varepsilon = 1$, in other words for a fixed choice of $\varepsilon \in \mathbb{R}_+$. For our study, a choice of $\varepsilon = \varepsilon(k)$, which is increasing in $k \in \mathbb{R}_+$, is necessary as discussed before. Indeed, we have $\frac{n\Psi^2}{\tau} = 1$ and $\frac{n\Psi}{\psi} = \frac{\sqrt{\sigma_Y n}}{d_n}$, and get

$$\mathbb{E}_{f_Y}^n\Big(\sup_{h\in B_k} \bar{\nu}_{h,1}^2 - 2(1+2\varepsilon)\sigma_Y\frac{\Delta_g(k)}{n}\Big)_+ \leq \frac{C}{n}\left(\sigma_Y\Delta_g(k)\exp^{-K_1\varepsilon} + \frac{\Delta_g(k)d_n^2}{n}e^{-K_2C_\varepsilon\sqrt{\varepsilon}\sqrt{\sigma_Y n}d_n^{-1}}\right).$$

Choosing now $\varepsilon = 4\lambda k^\rho / K_1$ we get by applying assumption **[G2]** for $k \geq k_g$

$$\Delta_g(k)\exp(-K_1\varepsilon) \leq C_g k^{-2\gamma}\exp(-2\lambda k^\rho),$$

which is summable over $\mathbb{N}$. Next for $k \geq k_g$ we get $C_\varepsilon\sqrt{\varepsilon} \geq \varepsilon/2$ and choosing $d_n := \sqrt{n\sigma_Y}2K_2/K_1$ leads to

$$-K_2 C_\varepsilon \sqrt{\varepsilon}\sqrt{\sigma_Y n} d_n^{-1} \geq \lambda k^\rho,$$

which implies

$$\mathbb{E}^n_{f_Y}\big(\sup_{h\in B_k}\overline{\nu}^2_{h,1} - 2(1+2\varepsilon)\sigma_Y\frac{\Delta_g(k)}{n}\big)_+ \leq \frac{C_g\sigma_Y}{n}2k^{-2\gamma}\exp(-2\lambda k^\rho).$$

Furthermore, for $\chi \geq \chi_0 > 0$ holds $\frac{1}{24}\mathrm{pen}(k) \geq 2(1+4\lambda k^\rho)/K_1)\sigma_Y\Delta_g(k)n^{-1}$ implying

$$\sum_{k\leq K_n}\mathbb{E}^n_{f_Y}(\sup_{t\in B_k}|\overline{\nu}_{h,1}|^2 - \frac{1}{24}\mathrm{pen}(k))_+ \leq \frac{C_g}{n}\sigma_Y\sum_{k\leq K_n}k^{-2\gamma}\exp(-2\lambda k) \leq \frac{C(g)\sigma_Y}{n}.$$

Now, we consider M_2. Let us define $\overline{f}_k := \widehat{f}_k - \widetilde{f}_k$. Then from $\overline{\nu}_{h,2} = \overline{\nu}_h - \overline{\nu}_{h,1}$ we deduce $\overline{\nu}^2_{h,2} = \langle\overline{f}_k - \mathbb{E}^n_{f_Y}(\overline{f}_k), h\rangle^2_{x^{2c-1}} \leq \|\overline{f}_k - \mathbb{E}^n_{f_Y}(\overline{f}_k)\|^2_{x^{2c-1}}$ for any $h\in B_k$. Further,

$$\max_{k\in\mathcal{K}_n}\|\overline{f}_k - \mathbb{E}^n_{f_Y}(\overline{f}_k)\|^2_{x^{2c-1}} \leq \|\overline{f}_{K_n} - \mathbb{E}^n_{f_Y}(\overline{f}_{K_n})\|^2_{x^{2c-1}}$$

and

$$\begin{aligned}\mathbb{E}^n_{f_Y}(\|\overline{f}_{K_n} - \mathbb{E}^n_{f_Y}(\overline{f}_{K_n}))\|^2_{x^{2c-1}}) &= \frac{1}{2\pi}\int_{[-K_n,K_n]}\frac{\mathbb{V}\mathrm{ar}^n_{f_Y}(\widehat{\mathcal{M}}(t) - \widetilde{\mathcal{M}}(t))}{\mathcal{M}_c[g](t)|^2}d\lambda(t)\\ &\leq \frac{\Delta_g(K_n)}{n}\mathbb{E}_{f_Y}(Y_1^{2(c-1)}\mathbb{1}_{(d_n,\infty)}(Y_1^{c-1}))\\ &\leq \mathbb{E}_{f_Y}(Y_1^{2(c-1)}\mathbb{1}_{(d_n,\infty)}(Y_1^{c-1})).\end{aligned}$$

Now $\mathbb{E}_{f_Y}(Y_1^{2(c-1)})\mathbb{1}_{(d_n,\infty)}(Y_1^{c-1})) \leq d_n^{-2}\mathbb{E}_{f_Y}(Y_1^{4(c-1)}) \leq C(\mathbb{E}_{f_Y}(Y_1^{4(c-1)})))n^{-1}$, which implies

$$\mathbb{E}^n_{f_Y}\left(\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}} - \frac{1}{12}\mathrm{pen}(\widehat{k})\right)_+ \leq \frac{C(g,\mathbb{E}_{f_Y}(Y_1^{5(c-1)})}{n}.$$

□

Before studying the resulting rates of convergence under regularity conditions we compare shortly the assumptions of Theorem 3.3.7 and Theorem 3.2.8.

Remark 3.3.13 (Continuation of Remark 3.3.11)**:**
Comparing the assumptions of Theorem 3.2.8 and Theorem 3.3.7, we observe that the boundness of $x \mapsto x^{2c-1}g(x)$ is not necessary in the case of super smooth error densities. More precisely, it is needed for the choice of $\tau := \|x^{2c-1}g\|_\infty \|G_k\|_\infty \mathbb{E}_f(X_1^{2(c-1)})$ of the Talagrand constant in the proof of Theorem 3.2.8. As mentioned in Remark 3.3.11 this choice cannot be applied in general in the super smooth error density case. Thus, an overestimation of the variance term is necessary, which on the other-hand does not need the assumption $\|x^{2c-1}g(x)\|_\infty < \infty$.
At the same time, the overestimation of the variance does effect the expectation $\mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - 2\widehat{\mathrm{pen}}(\widehat{k}))_+)$. Indeed, the moment assumption $\mathbb{E}_{f_Y}(Y_1^{5(c-1)}) < \infty$, compared to $\mathbb{E}_{f_Y}(Y_1^{4(c-1)}) < \infty$ in the proof of Theorem 3.2.8, allows us to bound the term with the order n^{-1}, that is

$$\mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - 2\widehat{\mathrm{pen}}(\widehat{k}))_+) \leq \frac{C(\chi, \mathbb{E}_{f_Y}(Y_1^{5(c-1)})}{n}.$$

Here, two things should be mentioned. First, the assumption $\mathbb{E}_{f_Y}(Y_1^{4(c-1)}) < \infty$ implies already in the super smooth case an order $\log(n)n^{-1}$,

$$\mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - 2\widehat{\mathrm{pen}}(\widehat{k}))_+) \leq \frac{C(\chi, \mathbb{E}_{f_Y}(Y_1^{4(c-1)}) \log(n)}{n}.$$

In other words a small digression of the rate by a Log-term occurs which could be seen as a negligible price to pay compared to a higher moment condition. Secondly, the moment assumption $\mathbb{E}_{f_Y}(Y_1^{5(c-1)}) < \infty$ can be weakened by an assumption $\mathbb{E}_{f_Y}(Y_1^{(4+\delta)(c-1)}) < \infty$ for $\delta \in (0,1)$, with

$$\mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - 2\widehat{\mathrm{pen}}(\widehat{k}))_+) \leq \frac{C(\delta, \chi, \mathbb{E}_{f_Y}(Y_1^{(4+\delta)(c-1)})}{n},$$

which was not considered for the sake of readability of the proof.

We will come to the deduction of the resulting rate under regularity assumptions of the density f for the fully data-driven estimator $\widehat{f}_{\widehat{k}}$.

Corollary 3.3.14:
Let $c \in \mathbb{R}$, $s, L \in \mathbb{R}_+$ and $f \in \mathbb{D}^s_c(L)$. Additionally assume that $\mathbb{E}_{f_Y}(Y_1^{5(c-1)}) < \infty$ and **[G2]** is fulfilled. Then for all $\chi > \chi_0$ holds

$$\mathbb{E}^n_{f_Y}(\|\widehat{f}_{\widehat{k}} - f\|^2_{x^{2c-1}}) \leq C(L, s, g, \mathbb{E}_f(Y_1^{5(c-1)}, \chi) \log(n)^{-2s/\rho},$$

where $C(L, s, g, \mathbb{E}_{f_Y}(Y_1^{5(c-1)}) > 0$ is a positive constant depending on L, s, g, χ and $\mathbb{E}_{f_Y}(Y_1^{5(c-1)})$.

Here, we deduce the same rate as the choice $k_o = (\log(n)/4\lambda)^{1/\rho}$ from Proposition 3.3.5. Contrary to the smooth case, this choice is independent of $s \in \mathbb{R}_+$ and thus known in advance. In this situation, we do not improve the rate by using a data-driven method. Again, this rate is rather pessimistic and in explicit situations, the data-driven method leads to a faster convergence rate. We consider now exemplary the case of a Log-Normal distributed random variable X. Its proof is a direct consequence and thus omitted.

Corollary 3.3.15:
Let $c \in \mathbb{R}$ such that $\mathbb{E}_g(U_1^{5(c-1)}) < \infty$, **[G2]** is fulfilled with $\rho = 1$ and $X \sim \mathrm{LN}_{(\mu,\sigma^2)}$. Then for all $\chi > \chi_0$ and $\varepsilon \in (0,1)$

$$\mathbb{E}^n_{f_Y}(\|\widehat{f}_{\widehat{k}} - f\|^2_{x^{2c-1}}) \leq C(g,\mu,\sigma,\chi\mathbb{E}_g(U_1^{5(c-1)}),\varepsilon)n^{-1+\varepsilon},$$

where $C(g,\mu,\sigma,\chi\mathbb{E}_g(U_1^{5(c-1)}),\varepsilon)$ is a positive constant depending on g μ, σ, $\mathbb{E}_g(U_1^{5(c-1)})$ and ε.

Thus, in the case of a Log-Normal distributed X_1, respectively if the decay of the Mellin transform of f is faster than the decay of $\mathcal{M}_g$, a better rate can be achieved through the data-driven choice $\widehat{k}$.

3.3.4 Numerical results

In this subsection, we compare the behavior of the data-driven spectral cut-off estimator $\widehat{f}_{\widehat{k}}$, presented in equation (3.14), with the data-driven choice (3.15) for the smooth error case and (3.18) in the super smooth case. To do so, we use the following examples for the unknown density f,

(i) Beta Distribution: $f(x) = B(2,5)^{-1}x(1-x)^4\mathbb{1}_{(0,1)}(x), x \in \mathbb{R}_+$ and

(ii) Log-Normal Distribution: $f(x) = (0.32\pi x^2)^{-1/2}\exp(-\log(x)^2/0.32)\mathbb{1}_{(0,\infty)}(x), x \in \mathbb{R}_+$.

A detailed discussion of these examples in terms of the decay of their Mellin transform can be found in section 2.5. To compare the behavior of the estimator in the smooth and the super smooth error case, we use the following examples of error densities g

a) Pareto Distribution: $g(x) = \mathbb{1}_{(1,\infty)}(x)x^{-2}$, $x \in \mathbb{R}_+$, and

b) Chi-squared Distribution: $g(x) = \mathbb{1}_{(0,\infty)}(x)\frac{x^{-1/2}}{\sqrt{2\pi}}\exp(-x/2)$, $x \in \mathbb{R}_+$, and

c) Gamma Distribution: $g(x) = \mathbb{1}_{(0,\infty)}(x)\frac{x^2}{2}\exp(-x)$, $x \in \mathbb{R}_+$.

Here it is worth pointing out that the example a) fulfill **[G1]** with $\gamma = 1$ and b) fulfills **[G2]** with $\rho = 1$. By minimizing an integrated weighted squared error over a family of histogram densities with randomly drawn partitions and weights we select $\chi = 7$ for a), and for the cases b) and c) we choose $\chi = 1.7$.
In Figure 3.7, we presented the choice $c = 1$ since for $c = 1/2$ in the case of b) we have no theoretical result of the behavior of $\widehat{f}_{\widehat{k}}$ as the assumptions of Theorem 3.2.8 hold not true. It can be seen that the reconstruction of the density f given the noisy sample $(Y_j)_{j\in[\![n]\!]}$ seems to be more difficult in the case of super smooth errors compared to smooth error which coincides with the theory.

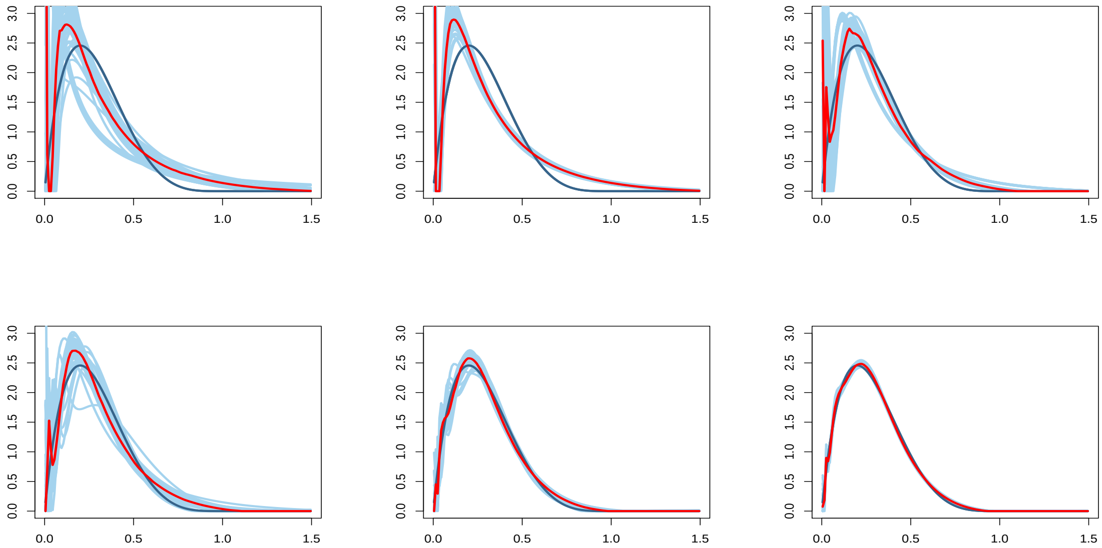

Figure 3.7: Considering the estimator $\widehat{f}_{\widehat{k}}$ and a sample size $n = 10^3$ (left), $n = 10^4$ (middle) and $n = 10^5$ the adaptive estimators are depicted for 50 Monte-Carlo simulations in the cases (i) with error density b) (top) and a) (bottom) and with $c = 1$. The true density f is given by the dark blue curve while the red curve is the point-wise empirical median of the 50 estimates.

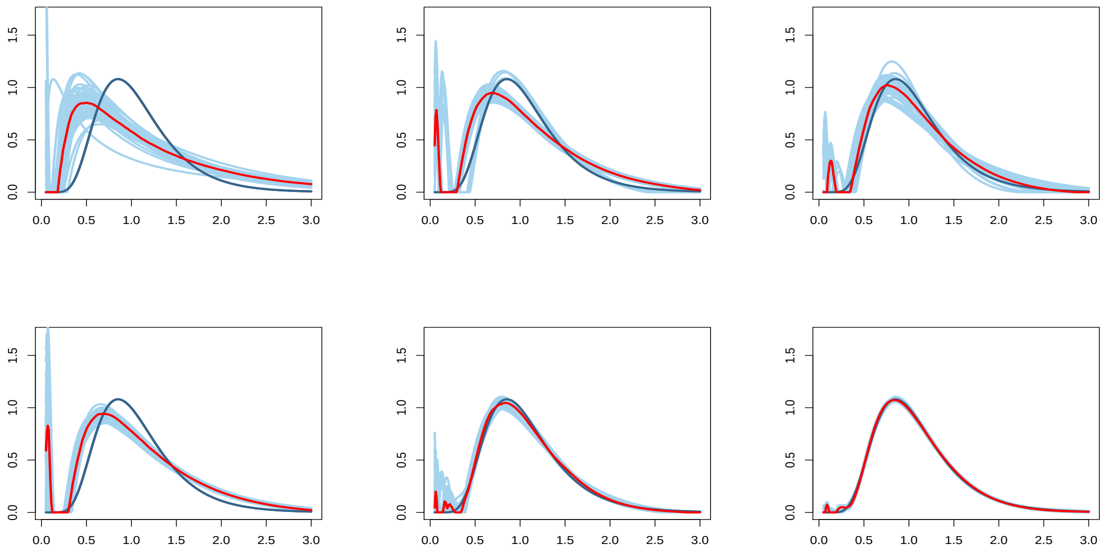

Figure 3.8: Considering the estimator $\widehat{f}_{\widehat{k}}$ and a sample size $n = 10^3$ (top) and $n = 10^5$(bottom) the adaptive estimators are depicted for 50 Monte-Carlo simulations in the cases (ii) with error density $b)$ (left), $c)$ (middle) and $a)$ (right) with $c = 1$. The true density f is given by the dark blue curve while the red curve is the point-wise empirical median of the 50 estimates.

Again in Figure 3.8, we study $c = 1$ since for $c = 1/2$ in the case of b) we have no theoretical result of the behavior of $f_{\widehat{k}}$ as the assumptions of Theorem 3.2.8 hold not true. Comparing the left and the middle column in Figure 3.8 with the right column, the different behavior of the smooth and the super smooth case is easy to be seen. It is interesting though that the middle column, the case of the Gamma-distributed error, the estimator seems to perform better. From a theoretical point of view, both left and middle column have Gamma-distributed errors, keeping in mind that the chi-squared distribution is just a special case of the Gamma distribution. This effect might come from the fact that the chi-squared distribution b) has the form parameter $p = 1/2$ while the considered Gamma distribution c) has the form parameter $p = 3$. Based on Example 3.3.2 the Mellin transform for the case b) decays faster than c). While this slight difference do not effect the general rate, it could effect the performance of the estimator for finite sample sizes.

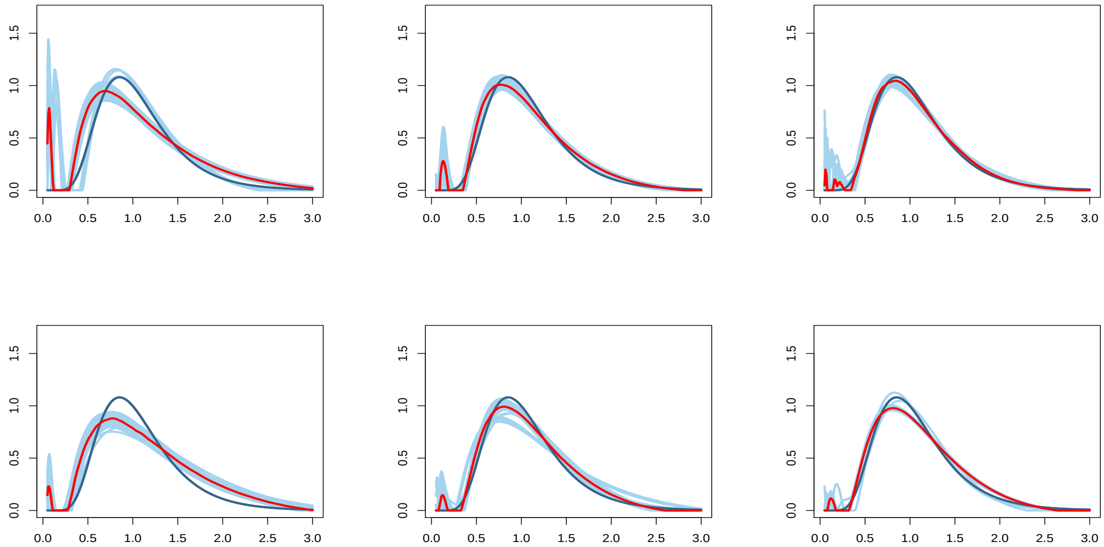

Figure 3.9: Considering the estimator $\widehat{f}_{\widehat{k}}$ and a sample size $n = 10^3$ (left), $n = 10^4$ (middle) and $n = 10^5$ (right) the adaptive estimators are depicted for 50 Monte-Carlo simulations in the cases (ii) with $c = 1$ (top) and $c = 1/2$ (bottom) and in the case of c). The true density f is given by the dark blue curve while the red curve is the point-wise empirical median of the 50 estimates.

In Figure 3.9, we compared the cases $c = 1$ with $c = 1/2$. It can be seen that the behavior of the estimators close to 0 differs strongly. This is natural since for the case $c = 1$ the weight function suppress the influence of this region on the global risk. This effect does not occur for $c = 1/2$.

3.3.5 Conclusion

In Section 3.3 we have studied the spectral cut-off estimator, presented in Section 3.2, for super smooth error densities. More precisely, we have derived an upper bound for the minimax risk and showed that the resulting rate is minimax-optimal for the case $c = 1$ for a variety of super smooth error densities. We then considered for the case $\rho \geq 1$ a fully data-driven estimator of the density, derived an oracle inequality, and showed that the estimator achieves the minimax rate. The finite sample size properties of our estimator has been motivated by a Monte-Carlo simulation study.

In general, we can say that the estimation of the density with super smooth error is possible but harder in terms of the decay of the risk. For the lower bound, a proof of the case $c \neq 1$ is left as an open research question. For the data-driven method the cases $\rho \in (0,1)$ can be considered, but have been omitted for the sake of readability. More precisely, for the upcoming theory, the case of $\rho = 1$ will be of special interest while considering the estimation of volatility densities in a stochastic volatility model, see Section 3.6.4.

3.3.6 Proofs

The next inequality was proven by Nagaev (1979). A similar formulation can be found in Liu et al. (2013), equation (1.3).

Lemma 3.3.16 (Nagaev's inequality):
Let $X_1, \dots, X_n$ be i.i.d. mean-zero random variables with $\mathbb{E}(|X_1|^p) < \infty$ for $p > 2$. Then for any $x \in \mathbb{R}_+$ holds

$$\mathbb{P}(\sum_{j \in [\![n]\!]} X_j \geq x) \leq (1 + 2p^{-1})^p \frac{n\mathbb{E}(|X_1|^p)}{x^p} + 2\exp(-\frac{a_p x^2}{n\mathbb{E}(X_1^2)}),$$

where $a_p = 2e^{-p}(p+2)^{-2}$.

Proof of Lemma 3.3.4. Let us define

$$\varphi_1(y) := \int_0^y (1+t^2)^{-\gamma} \exp(2\lambda t^\rho) dt \text{ and } \varphi_2(y) := y^{1-2\gamma-\rho} \exp(2\lambda y^\rho)$$

for $y \in \mathbb{R}_+$. Then it is sufficient to show that we can find a constant $C_{\gamma,\lambda,\rho} \in \mathbb{R}_+$ such that $\varphi_1(y) \leq C_{\gamma,\lambda,\rho}\varphi_2(y)$ for all $y > y_{\gamma,\lambda,\rho} \in \mathbb{R}_+$. Indeed, since both functions are strictly positive and continuous, we can always find for bounded subintervals $(0, y_{\gamma,\lambda,\rho}) \subset \mathbb{R}_+$ a constant $C_{\gamma,\lambda,\rho}$m such that the inequality holds. Therefore, we are interested in the behavior of φ_1/φ_2 for large arguments. Using the L'Hôpital rule we get

$$\begin{aligned}\lim_{t\to\infty} \frac{\varphi_1(t)}{\varphi_2(t)} &= \lim_{t\to\infty} \frac{(1+t^2)^{-\gamma}\exp(2\lambda t^\rho)}{(1-2\gamma-\rho)t^{-2\gamma-\rho}\exp(2\lambda t^\rho) + 2\lambda\rho t^{-2\gamma}\exp(2\lambda t^\rho)} \\ &= \lim_{t\to\infty} \frac{(1+t^{-2})^{-\gamma}}{(1-2\gamma-\rho)t^{-\rho} + 2\lambda\rho} = (2\lambda\rho)^{-1}.\end{aligned}$$

Thus, we can find a $y_{\rho,\lambda,\gamma} > 0$ such that $\varphi_1(y)/\varphi_2(y) \leq (\lambda\rho)^{-1}$ for all $y > y_{\rho,\lambda,\gamma}$. This implies the claim. □

Although the result of Lemma 3.2.9 and Lemma 3.3.8 seem to be similar, we present an alternative proof of the decomposition in the proof of Lemma 3.3.8.

Proof of Lemma 3.3.8. Let $k \in \mathcal{K}_n$ and let us keep in mind that $[-k', k'] = \operatorname{supp}(\mathcal{M}_c[f_{k'}])$, for $k' \in \mathcal{K}_n$. For $K_n := \lfloor \log(n)^{1/\rho} \rfloor$, for all $k' \in \mathcal{K}_n$ we have $[-k', k] \subseteq [-K_n, K_n]$. Further, we have for any $k' \in \mathcal{K}_n$ that $\|\widehat{f}_{K_n}\|^2_{x^{2c-1}} - \|\widehat{f}_{k'}\|^2_{x^{2c-1}} = \|\widehat{f}_{K_n} - \widehat{f}_{k'}\|^2_{x^{2c-1}}$ implying with (3.18)

$$\|\widehat{f}_{\widehat{k}} - \widehat{f}_{K_n}\|^2_{x^{2c-1}} + \widehat{\operatorname{pen}}(\widehat{k}) \leq \|\widehat{f}_k - \widehat{f}_{K_n}\|^2_{x^{2c-1}} + \widehat{\operatorname{pen}}(k).$$

Now for every $k' \in \mathcal{K}_n$ we have

$$\|\widehat{f}_{k'} - f_{K_n}\|^2_{x^{2c-1}} = \|\widehat{f}_{k'} - \widehat{f}_{K_n}\|^2_{x^{2c-1}} + \|\widehat{f}_{K_n} - f_{K_n}\|^2_{x^{2c-1}} + 2\langle \widehat{f}_{k'} - \widehat{f}_{K_n}, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}},$$

which implies that

$$\begin{aligned}\|\widehat{f}_{\widehat{k}} - f_{K_n}\|^2_{x^{2c-1}} - \|\widehat{f}_k - f_{K_n}\|^2_{x^{2c-1}} &= \|\widehat{f}_{\widehat{k}} - \widehat{f}_{K_n}\|^2_{x^{2c-1}} - \|\widehat{f}_k - \widehat{f}_{K_n}\|^2_{x^{2c-1}} + 2\langle \widehat{f}_{\widehat{k}} - \widehat{f}_k, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}} \\ &\leq \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k}) + 2\langle \widehat{f}_{\widehat{k}} - \widehat{f}_k, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}}, \end{aligned} \tag{3.20}$$

As

$$\begin{aligned}\langle \widehat{f}_{\widehat{k}} - \widehat{f}_k, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}} &= \langle (\widehat{f}_{\widehat{k}} - f_{\widehat{k}}) + (f_{\widehat{k}} - f_k) + (f_k - \widehat{f}_k), \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}} \\ &= \|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}} + \langle f_{\widehat{k}} - f_k, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}} - \|\widehat{f}_k - f_k\|^2_{x^{2c-1}}. \end{aligned} \tag{3.21}$$

Now combining (3.20) and (3.21), we get

$$\begin{aligned}\|\widehat{f}_{\widehat{k}} - f_{K_n}\|^2_{x^{2c-1}} \leq{}& \|f_k - f_{K_n}\|^2_{x^{2c-1}} - \|\widehat{f}_k - f_k\|^2_{x^{2c-1}} + 2\langle f_{\widehat{k}} - f_k, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}} \\ &+ \widehat{\mathrm{pen}}(k) + 2\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}} - \widehat{\mathrm{pen}}(\widehat{k}). \end{aligned} \tag{3.22}$$

Let us consider the term $|2\langle f_{\widehat{k}} - f_k, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}}|$. First we remind that for any $k' \in \mathcal{K}_n$

$$\|\widehat{f}_{k'} - f_{k'}\|^2_{x^{2c-1}} = \frac{1}{2\pi}\int_{\mathbb{R}} \mathbb{1}_{[-k',k']}(t)\frac{|\mathcal{M}_c[f_Y](t) - \widehat{\mathcal{M}}_c(t)|^2}{|\mathcal{M}_c[g](t)|^2}d\lambda(t).$$

Setting $A^* := [-\widehat{k}, \widehat{k}] \cup [-k, k]$, we have $\mathcal{M}_c[f_{\widehat{k}} - f_k] = \mathcal{M}_c[f](\mathbb{1}_{[-\widehat{k},\widehat{k}]} - \mathbb{1}_{[-k,k]})$ implying that $\mathrm{supp}(\mathcal{M}_c[f_{\widehat{k}} - f_k]) \subseteq A^* \subseteq [-K_n, K_n]$ by definition of K_n. Using the Cauchy-Schwarz inequality and $2ab \leq a^2 + b^2$ we deduce

$$\begin{aligned}|2\langle f_{\widehat{k}} - f_k, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}}| &= \frac{2}{(2\pi)^2}\left|\int_{A^*} \mathcal{M}_c[f_{\widehat{k}} - f_k](t)\frac{\widehat{\mathcal{M}}_c(-t) - \mathcal{M}_c[f_Y](-t)}{\mathcal{M}_c[g](-t)}d\lambda(t)\right| \\ &\leq \frac{1}{4}\|f_{\widehat{k}} - f_k\|^2_{x^{2c-1}} + \frac{4}{2\pi}\int_{\mathbb{R}} \mathbb{1}_{A^*}(t)\frac{|\mathcal{M}_c[f_Y](t) - \widehat{\mathcal{M}}_c(t)|^2}{|\mathcal{M}_c[g](t)|^2}d\lambda(t) \\ &\leq \frac{1}{2}\|f_{\widehat{k}} - f_{K_n}\|^2_{x^{2c-1}} + \frac{1}{2}\|f_k - f_{K_n}\|_{x^{2c-1}} + 4\|\widehat{f}_k - f_k\|^2_{x^{2c-1}} + 4\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}} \end{aligned}$$

using that $\mathbb{1}_{A^*} \leq \mathbb{1}_{A_k} + \mathbb{1}_{A_{\widehat{k}}}$. Thus, we get

$$\begin{aligned}&|2\langle f_{\widehat{k}} - f_k, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}}| \\ &\qquad\leq \frac{1}{2}\|f_{K_n} - f_k\|^2_{x^{2c-1}} + \frac{1}{2}\|\widehat{f}_{\widehat{k}} - f_{K_n}\|^2_{x^{2c-1}} + 4\|\widehat{f}_k - f_k\|^2_{x^{2c-1}} + \frac{7}{2}\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}}, \end{aligned}$$

which implies with (3.52)

$$\|\widehat{f}_{\widehat{k}} - f_{K_n}\|^2_{x^{2c-1}} \leq 3\|f_k - f_{K_n}\|^2_{x^{2c-1}} + 6\|\widehat{f}_k - f_k\|^2_{x^{2c-1}} + 2\widehat{\mathrm{pen}}(k) + 11\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}} - 2\widehat{\mathrm{pen}}(\widehat{k}).$$

Since $\mathbb{E}^n_{f_Y}(\widehat{\mathrm{pen}}(k)) = \mathrm{pen}(k)$ and for $\chi_0 \geq 6$ $6\mathbb{E}^n_{f_Y}(\|\widehat{f}_k - f_k\|^2_{x^{2c-1}}) \leq 6\sigma_Y\Delta_g(k)n^{-1} \leq \mathrm{pen}(k)$. we get

$$\begin{aligned}\mathbb{E}^n_{f_Y}(\|\widehat{f}_{\widehat{k}} - f_{K_n}\|^2_{x^{2c-1}}) \leq{}& 3\left(\|f_{K_n} - f_k\|^2_{x^{2c-1}} + \mathrm{pen}(k)\right) + 11\mathbb{E}^n_{f_Y}\left(\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}} - \frac{1}{12}\mathrm{pen}(\widehat{k})\right)_+ \\ &+ \mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - 2\widehat{\mathrm{pen}}(\widehat{k}))_+), \end{aligned}$$

which implies the claim since $\|\widehat{f}_{\widehat{k}} - f\|^2_{x^{2c-1}} = \|\widehat{f}_{\widehat{k}} - f_{K_n}\|^2_{x^{2c-1}} + \|f - f_{K_n}\|^2_{x^{2c-1}}$. □

Proof of Lemma 3.3.10. Let us define $\Omega := \{|\widehat{\sigma}_Y - \sigma_Y| \leq \sigma_Y/2\}$. Then on Ω we have $2\widehat{\sigma}_Y \geq \sigma_Y$ and

$$\begin{aligned}
\mathbb{E}^n_{f_Y}((\mathrm{pen}(k) - 2\widehat{\mathrm{pen}}(k))_+) &= \chi \mathbb{E}^n_{f_Y}\left(\widehat{k}\frac{\Delta_g(\widehat{k}^\rho)}{n}(\sigma_Y - 2\widehat{\sigma}_Y)_+\right) \\
&\leq 2\chi K_n^\rho \mathbb{E}^n_{f_Y}(|\sigma_Y - \widehat{\sigma}_Y| \mathbb{1}_{\Omega^c}) \\
&\leq C(\chi, g)\log(n)\frac{\sqrt{\mathbb{V}\mathrm{ar}^n_{f_Y}(\widehat{\sigma}_Y)\mathbb{P}(\Omega^c)}}{\sigma_Y}.
\end{aligned}$$

Now applying the Nagaev inequality, Lemma 3.3.16, we get for $p = 5$

$$\mathbb{E}^n_{f_Y}((\mathrm{pen}(k) - 2\widehat{\mathrm{pen}}(k))_+) \leq C(\chi, \mathbb{E}_{f_Y}(Y_1^{5(c-1)}))\frac{\log(n)}{n^{-5/2}} \leq \frac{C(\chi, \mathbb{E}^n_{f_Y}(Y_1^{5(c-1)}))}{n}.$$

□

3.4 Under multiplicative measurement errors with oscillating densities

3.4.1 Introduction

As mentioned in Section 3.1, from an inverse problem theoretical point of view, the estimator presented in Equation (3.14) can be interpreted as a spectral cut-off regularization of the inverse Mellin transform to ensure that the function $\widehat{\mathcal{M}}_c/\mathcal{M}_c[g]$ lies in its domain, respectively introduced with $\mathcal{M}_c^+[g] := \mathbb{1}_{(-k,k)}/\mathcal{M}_c[g]$ a regularized version of $1/\mathcal{M}_c[g]$ such that $\widehat{\mathcal{M}}_c\mathcal{M}_c^+[g]$ lies in the domain of $\mathcal{M}_c^\dagger$. In this section, we will propose another choice of $\mathcal{M}_c^+[g]$ based on a so-called Ridge-approach similar to the work of Meister (2009) and Hall and Meister (2007).

Despite this, we stay in the global density estimation under the multiplicative measurement error model. As a reminder, we are interested in estimating the density $f : \mathbb{R}_+ \to \mathbb{R}_+$ of a positive random variable X given an i.i.d sample $(Y_j)_{j\in[\![n]\!]}$ of

$$Y = XU,$$

where U is a second positive random variable, stochastically independent of X, with known error density $g : \mathbb{R}_+ \to \mathbb{R}_+$.

3.4.2 Estimation strategy

Given a positive, continuous Ridge function $\rho_n : \mathbb{R} \to \mathbb{R}_+$, possibly dependent on the sample size $n \in \mathbb{N}$, we can state, with $\max(|\mathcal{M}_c[g](t)|, \rho_n(t)) =: (|\mathcal{M}_c[g]| \vee \rho_n)(t) > 0$, for any $t \in \mathbb{R}$

$$\frac{|\mathcal{M}_c[g]|}{(|\mathcal{M}_c[g]| \vee \rho_n)^2} \leq \frac{|\mathcal{M}_c[g]|}{\rho_n^2}.$$

If $\rho_n^{-2} \in \mathbb{L}^2(\mathbb{R})$ then $\widehat{\mathcal{M}}_c \frac{\overline{\mathcal{M}_c[g]}}{|\mathcal{M}_c[g]|\vee\rho_n)^2} \in \mathbb{L}^2(\mathbb{R})$ follows. Here, $\widehat{\mathcal{M}}_c(t) := n^{-1}\sum_{j\in[\![n]\!]} Y_j^{c-1+it}$ is the empirical Mellin transform introduced in Section 3.1.2, respectively in Section 3.2.

For any ridge function $\rho_n : \mathbb{R} \to \mathbb{R}_+$ with $\rho_n^{-2} \in \mathbb{L}^2(\mathbb{R})$ and any error density $g : \mathbb{R}_+ \to \mathbb{R}_+$ with $\mathbb{E}(U_1^{c-1}) < \infty$ the Ridge estimator $\widetilde{f}_{\rho_n}$ is defined by

$$\widetilde{f}_{\rho_n} := \mathcal{M}_c^\dagger\left[\frac{\widehat{\mathcal{M}}_c\overline{\mathcal{M}_c[g]}}{(|\mathcal{M}_c[g]| \vee \rho_n)^2}\right].$$

If the function $\mathcal{M}_c[g]/\rho_n^2 \in \mathbb{L}^1(\mathbb{R})$ we can express this estimator explicitly by

$$\widetilde{f}_{\rho_n}(x) := \frac{1}{2\pi}\int_{\mathbb{R}} x^{-c-it}\widehat{\mathcal{M}}_c(t)\frac{\mathcal{M}_c[g](-t)}{((|\mathcal{M}_c[g]| \vee \rho_n)(t))^2}d\lambda(t). \tag{3.23}$$

Remark 3.4.1 (Regularized inversion)**:**
In comparison with the family of estimators $(\widehat{f}_k)_{k\in\mathbb{R}_+}$ from Section 3.2.2, both estimators are based on the application of the inverse Mellin transform $\mathcal{M}_c^\dagger$ on the product of the empirical Mellin transform $\widehat{\mathcal{M}}_c$ and a regularized version $\mathcal{M}_c^+[g]$ of $1/\mathcal{M}_c[g]$, which ensures that $\widehat{\mathcal{M}}_c\mathcal{M}_c^+[g] \in \mathbb{L}^1(\mathbb{R})\cap\mathbb{L}^2(\mathbb{R})$. In other words they follow the same construction

$$\widehat{f} = \mathcal{M}_c^\dagger[\widehat{\mathcal{M}}_c\mathcal{M}_c^+[g]], \qquad \text{where } \mathcal{M}_c^+[g] = \begin{cases} \frac{\mathbb{1}_{(-k,k)}}{\mathcal{M}_c[g]}, & \widehat{f} = \widehat{f}_k, \\ \frac{\overline{\mathcal{M}_c[g]}}{(|\mathcal{M}_c[g]|\vee\rho_n)^2}, & \widehat{f} = \widetilde{f}_{\rho_n}. \end{cases}$$

The regularization of the inverse operator to construct estimator is a frequently used technique in the theory of statistical inverse problems.

Before studying the properties of the estimator $\widetilde{f}_{\rho_n}$ and deducing situations where it is well-defined while its counterpart $\widehat{f}_k$ is undefined, we will begin by exploring on which subset of $\mathbb{R}$ the two choices of $\mathcal{M}_c^+[g]$ differ.
Let us define the set $G_n := \{t\in\mathbb{R} : \rho_n(t) > |\mathcal{M}_c[g](t)|\}$. Then for all $t \in G_n^c$, we can ensure that $|\mathcal{M}_c[g](t)| \geq \rho_n(t) > 0$ and thus

$$\widehat{\mathcal{M}}_c(t)\frac{\mathcal{M}_c[g](-t)}{(|\mathcal{M}_c[g]|\vee\rho_n)(t))^2} = \frac{\widehat{\mathcal{M}}_c(t)\mathcal{M}_c[g](-t)}{|\mathcal{M}_c[g](t)|^2} = \frac{\widehat{\mathcal{M}}_c(t)}{\mathcal{M}_c[g](t)}, \qquad \text{for any } t\in G_n^c,$$

where all quotients are well-defined even without the assumption **[G0]** from Section 3.2.

Remark 3.4.2 (Choices of ρ_n)**:**
It seems natural to consider a choice of ρ_n such that both $\rho_n^{-2}\in\mathbb{L}^2(\mathbb{R})$ and that ρ_n converges pointwise to 0, that is for any $t\in\mathbb{R}$ should $\rho_n(t)\to 0$ for $n\to\infty$. A choice of such a ρ_n is presented in Meister (2009) by

$$\rho_n(t) := n^{-\eta}(1+|t|)^\xi, \quad \xi,\eta\geq 0. \tag{3.24}$$

To ensure that $\rho_n^{-2}\in\mathbb{L}^2(\mathbb{R})$, it is sufficient to choose $\xi > 1/4$. In the work of Hall and Meister (2007), a notion of Ridge estimator is presented, which includes the choice of $\xi = 0$, that is ρ_n is a constant function. The square-integrablilty of the function $|\mathcal{M}_c[g]|/\rho_n^2$ can then only be achieved, if the function $\mathcal{M}_c[g]$ decays fast enough. This excludes many error densities, for example if g is Log-Gamma distributed with $a\leq 1/2$, compare Example 2.1.10. To include these cases, we need to modify our estimator in the following way inspired by the work of Hall and Meister (2007).

Let ρ_n be of the form as seen in Equation (3.24) and let $r\geq 0$ be such that $\mathcal{M}_c[g]^{r+1}\rho_n^{-r-2}\in\mathbb{L}^1(\mathbb{R})\cap\mathbb{L}^2(\mathbb{R})$. Then,

$$\mathrm{R}_{\rho_n,r,g} := \overline{\mathcal{M}_c[g]}|\mathcal{M}_c[g]|^r(|\mathcal{M}_c[g]|\vee\rho_n)^{-r-2}\in\mathbb{L}^1(\mathbb{R})\cap\mathbb{L}^2(\mathbb{R})$$

and we define the estimator $\widehat{f}_{\rho_n,r}$ for any $x\in\mathbb{R}_+$ by

$$\widehat{f}_{\rho_n,r}(x) := \frac{1}{2\pi}\int_{\mathbb{R}} x^{-c-it}\widehat{\mathcal{M}}_c(t)\mathrm{R}_{\rho_n,r,g}(t)d\lambda(t), \tag{3.25}$$

Using the usual bias-variance decomposition we get

$$\mathbb{E}^n_{f_Y}(\|f - \widehat{f}_{\rho_n,r}\|^2_{x^{2c-1}}) = \|f - \mathbb{E}^n_{f_Y}(\widehat{f}_{\rho_n,r})\|^2_{x^{2c-1}} + \frac{1}{2\pi}\int_{\mathbb{R}} \mathbb{V}\mathrm{ar}^n_{f_Y}(\widehat{\mathcal{M}}_c(t))|\mathrm{R}_{\rho_n,r,g}(t)|^2 d\lambda(t).$$

Bounding both the bias and the variance term separately we get the following results whose proof is postponed to Section 3.2.7.

Proposition 3.4.3 (Ridge Upper bound)**:**
Let $c \in \mathbb{R}$ and $f \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$. Let ρ_n as in (3.24) and $r \geq 0$ such that $\mathcal{M}_c[g]^{r+1}/\rho_n^{r+2} \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ and $\sigma_c := \mathbb{E}(Y_1^{2(c-1)}) < \infty$. Then

$$\mathbb{E}(\|f - \widehat{f}_{\rho_n,r}\|^2_{x^{2c-1}}) \leq \|\mathbb{1}_{G_n}\mathcal{M}_c[f]\|^2_{\mathbb{R}} + \frac{\sigma_c}{2\pi n}\|\mathrm{R}_{\rho_n,r,g}\|^2_{\mathbb{R}},$$

where $G_n := \{t \in \mathbb{R} : \rho_n(t) > |\mathcal{M}_c[g](t)|\}$ and $\widehat{f}_{\rho_n,r}$ is defined in Equation 3.25.

For the Ridge function ρ_n we already stated that it is desirable that it converges point-wise to 0. In fact, this property would imply that $(G_n)_{n\in\mathbb{N}}$ is a decreasing sequence of sets, that ist $G_{n+1} \subset G_n$, which converges in many cases to Lebesgue null-set. A case, where such a converges is not given, would be if the error density g vanishes on set $A \subseteq \mathbb{R}$ with strictly positive Lebesgue-measure $\lambda(A) > 0$. We want to stress out that this discussion is necessary at this point, since the Ridge estimator is well-defined even for densities not fulfilling **[G0]**. Nevertheless, the result of Proposition 3.4.3 does not already imply the consistency of the ridge estimator. This is due to the fact, that it is not clear that the squared bias term $(\|\mathbb{1}_{G_n}\mathcal{M}_c[f]\|^2_{\mathbb{R}})_{n\in\mathbb{N}}$ is vanishing. We will now give a minimal assumption to ensure, that we can define a consistent estimator using the Ridge estimator framework. We will from now on assume, that the Mellin transform of g is almost everywhere not 0, that is

$$\lambda(\{t \in \mathbb{R} : \mathcal{M}_c[g](t) = 0\}) = 0. \qquad \textbf{([G-1])}$$

Under the assumption **[G-1]** we can use the dominated convergence theorem to state that the first and the second term in Proposition 3.4.3 are converging to 0. From this we can instantly deduce the following result whose proof is therefore omitted.

Corollary 3.4.4 (Consistency)**:**
Let the assumption of Proposition 3.4.3 be fulfilled and **[G-1]** holds true. If $n^{-1}\|\mathrm{R}_{\rho_n,r,g}\|^2_{\mathbb{R}} \to 0$ for $n \to 0$, then

$$\mathbb{E}^n_{f_Y}(\|\widehat{f}_{\rho_n,r} - f\|^2_{x^{2c-1}}) \to 0,$$

for $n \to 0$ and $\widehat{f}_{\rho_n,r}$ defined in Equation (3.25).

Although the assumptions in Corollary 3.4.4 seem to be rather technical, we will see that they can be simplified while considering more precise classes of error densities, compare Subsection 3.2.3. Before we define this family of error densities let us shortly comment on the consistency of the presented estimator.

Remark 3.4.5 (Strong consistency):
In Proposition 3.4.3 we have seen that we can determine a set of assumptions which ensures by application of the Markov inequality, that $\|\widehat{f}_{\rho_n,r} - f\|^2_{x^{2c-1}} \to 0$ in probability. Here, we needed the additional assumption that $f \in \mathbb{L}^2(\mathbb{R}, x^{2c-1})$ to construct the estimator and show its properties. A less restrictive metric which can be considered would be the $\mathbb{L}^1(\mathbb{R}_+, x^0)$-metric, since for any density $f \in \mathbb{L}^1(\mathbb{R}_+, x^0)$ holds. Further, the Mellin transform developed in $c = 1$ is well-defined for any density f. In the book of Meister (2009) they proposed an estimator $\widehat{f}_V$ of the density $f_V : \mathbb{R} \to \mathbb{R}$ of a real random variable V given i.i.d. copies of $Z = V + \varepsilon$ where V and ε are stochastically independent. They were able to show that their estimator $\widehat{f}$ is strongly consistent in the $\mathbb{L}^1(\mathbb{R})$-sense, that is $\|\widehat{f}_V - f_V\|_{\mathbb{L}^1(\mathbb{R})} \to 0$ almost surely. Given the log transformed data, $\log(Y) = \log(X) + \log(U)$, we can use the estimator $\widehat{f}_V$ for $V = \log(X)$ and deduce the estimator $\widehat{f}_X(x) := \widehat{f}_V(\log(x))x^{-1}$ for any $x \in \mathbb{R}_+$. Then $\|\widehat{f}_X - f\|_{\mathbb{L}^1(\mathbb{R}_+,x^0)} = \|\widehat{f}_V - f_V\|_{\mathbb{L}^1(\mathbb{R})}$ implying that the estimator $\widehat{f}_X$ is strongly consistent in the $\mathbb{L}^1(\mathbb{R}_+, x^0)$ sense. Although it might be tempting generalize this result for the $\mathbb{L}^1(\mathbb{R}_+, x^{c-1})$-distance for any $c \in \mathbb{R}$, it would need an additional moment assumption on f which contradicts the idea of considering the most general case.

Oscillating error densities Examples of error densities which do not fulfill **[G0]** are given by the so-called OSCILLATING ERROR DENSITIES which we define analogously to the additive deconvolution in Hall and Meister (2007).

Definition 3.4.6 (Oscillating error density):
Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\mathbb{E}_g(U^{c-1}) < \infty$ for a $c \in \mathbb{R}_+$. Then we call g a OSCILLATING ERROR DENSITY if there exist $c_g, C_g, \lambda, T > 0, \mu, \rho \in \mathbb{R}_+, \gamma, \kappa \in \mathbb{R}$ and $\mu \geq 1$ such that

$$c_g(1 + |t|^2)^{-\gamma/2}|\sin(\lambda t)|^\mu e^{-\kappa|t|^\rho} \leq |\mathcal{M}_c[g](t)| \leq C_g(1 + |t|^2)^{-\gamma/2}|\sin(\lambda t)|^\mu e^{-\kappa|t|^\rho} \text{ for all } |t| > T$$

and $|\mathcal{M}_c[g](t)| > c_g$ for all $t \in [-T, T]$.

Definition 3.4.6 implies that for an oscillating error density the set of zeros $\{t \in \mathbb{R} : \mathcal{M}_c[g](t) = 0\}$ is countable. Since $\lim_{x\to 0} \sin(x)/x = 1$, respectively $\lim_{x\to x_o} \sin(x_o - x)/x_o - x = 1$ for any zero $x_0 \in \mathbb{R}$ of the sine function, we see that the zeros of the sin function are of order 1, implying, that the zeros $\{t \in \mathbb{R} : \mathcal{M}_c[g](t) = 0\}$ are of order $\mu \geq 1$. Therefore, the function $\mathbb{1}_{[-k,k]}/\mathcal{M}_c[g]$ is not square-integrable and thus **[G0]** is not fulfilled. We will now present an example of an oscillating error density.

Example 3.4.7 (Log-Uniform distribution):
Let $U \sim \mathrm{LU}_{(-1,1)}$, that is $\log(U) \sim \mathrm{U}_{(-1,1)}$. Then the density g of U is given by $g(x) = \frac{1}{2x}\mathbb{1}_{(e^{-1},e)}(x), \quad x \in \mathbb{R}_+$. Since $\mathbb{E}_g(U_1^{c-1}) < \infty$ for any $c \in \mathbb{R}$, its Mellin transform is well-defined for $t \in \mathbb{R}$ by

$$\mathcal{M}_c[g](t) = \frac{\sinh(c - 1 + it)}{c - 1 + it},$$

where $\sinh(z) := (\exp(z) - \exp(-z))/2$, $z \in \mathbb{C}$, is the sinus hyperbolicus. For $c = 1$ we have $\sinh(it) = i\sin(t), t \in \mathbb{R}$, and thus $\mathcal{M}_1[g](t) = \frac{\sin(t)}{t}, \quad t \in \mathbb{R}$, showing that g is for $c = 1$ an oscillating error density.

Considering Example 3.4.7, the natural question arise if $c = 1$ is the only value of $c \in \mathbb{R}$ such that $\mathcal{M}_c[g]$ is oscillating. Indeed, applying the calculation rule $\sinh(s + it) = \sinh(s)\cos(t) + i\cosh(s)\sin(t)$, for $s, t \in \mathbb{R}$, where $\cosh(z) := (\exp(z) + \exp(-z))/2$, $z \in \mathbb{C}$ to get

$$\mathcal{M}_c[g](t) = 0 \Leftrightarrow (\sinh(c-1) = 0) \vee (\cosh(c-1) = 0) \Leftrightarrow c = 1,$$

since $\cosh(x) > 0$ for $x \in \mathbb{R}$ and $x = 0$ is the only zero of sinh on $\mathbb{R}$. In other words, the $\mathrm{LU}_{(-1,1)}$ distribution is only for $c = 1$ an oscillating error density. In the additive noise model, similar results motivate the usage of the bilateral Laplace transform, instead of the Fourier transform, as pointed out in Belomestny and Goldenshluger (2021).

Remark 3.4.8 (Existence of a zero-free Mellin transform)**:**
From Example 3.4.7 and the further discussion, one might wonder if there always exist a value of $c \in \mathbb{R}$ such that the Mellin transform $\mathcal{M}_c[g]$ of g is zero-free. Indeed, if there exists a $c' \in \mathbb{R} \setminus \{1\}$ such that $\mathbb{E}_g(U^{c'-1}) < \infty$, then the Mellin transform of g is well-defined for all $c \in (c', 1)$, respectively $(1, c')$. Now from the interpretation

$$\mathcal{M}[g] : \{z \in \mathbb{C} : \mathrm{Re}(z) \in (c', 1)\} \to \mathbb{C}, z = s + it \mapsto \mathcal{M}[g](z) := \mathcal{M}_s[g](t),$$

it is known from the complex analysis literature, that the function $\mathcal{M}[g]$ is holomorph on $\{z \in \mathbb{C} : \mathrm{Re}(z) \in (c', 1)\}$, compare Paris and Kaminski (2001). From this we deduce that there exists a $c \in (c', 1)$ such that $\mathcal{M}_c[g]$ is zero-free. Otherwise, the holomorphic function $\mathcal{M}[g]$ would have an uncountable set of zero which implies that $\mathcal{M}[g]$ is constant zero, due to the identity theorem of holomorphic functions. The only distribution with a constant 0 Mellin transform is the point measure in 0.

Remark 3.4.9 (Counter example for a zero-free Mellin transform)**:**
On the other hand, not all distributions possesses a $c \in \mathbb{R}_+ \setminus \{1\}$ with $\mathbb{E}_g(U^{c-1}) < \infty$. For example, considering the Log-Cauchy distribution, with density

$$g_V(u) := \frac{1}{x\pi(1 + |\log(u)|^2)}, \quad u \in \mathbb{R}_+,$$

fulfills $\mathbb{E}_{g_V}(V^{c-1}) < \infty$ only for $c = 1$.
Therefore the multiplicative convolution of a Log-Cauchy and a Log-Uniform distribution would be an example of an oscillating error density without the existence of $c \in \mathbb{R}$ such that $\mathcal{M}_c[g]$ is zero-free.
More precisely, let $U = VW$, where $V \sim g_V$ and $W \sim g_W(x) := (2x)^{-1}\mathbb{1}_{(e^{-1},e)}(x)$, $x \in \mathbb{R}_+$. If additionally V and W are stochastically independent, then $\mathbb{E}(U^{c-1}) < \infty$ if and only if $c = 1$ and for the density g of U holds

$$\mathcal{M}_1[g](t) = \mathcal{M}_1[g_V](t)\mathcal{M}_1[g_W](t) = \frac{\sin(t)}{t}\exp(-|t|),$$

which is an example for a super smooth oscillating error density.

In the next section, we will show that the proposed estimator is minimax-optimal for smooth error densities and comment on the case of oscillating error densities.

3.4.3 Minimax theory

For the further analysis, we consider the case of a constant ridge function, that is $\rho_n(t) := n^{-\eta}$ which correspond to $\xi = 0$, compare Equation (3.24). We can now prove that in the case of a smooth error densities the ridge estimator does reach the minimax rate developed in Section 3.2.3. As a reminder, we call an error density smooth if there exist constants $c_g, C_g, \gamma \in \mathbb{R}_+$ such that

$$\forall t \in \mathbb{R} : c_g(1+|t|^2)^{-\gamma/2} \leq |\mathcal{M}_c[g](t)| \leq C_g(1+|t|^2)^{-\gamma/2}. \qquad \textbf{([G1])}$$

Under assumption **[G1]** it is sufficient to choose $r > 0 \vee (\gamma^{-1} - 1)$ to ensure that $\mathcal{M}_c[g]^{r+1}/\rho_n^{r+2} \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. Now let us show that for a particular choice of $\eta \in \mathbb{R}_+$ the resulting ridge estimator achieves the minimax rate over the Mellin Sobolev ellipsoids introduced in Section 3.1. The proof of the following Proposition is rather technical and thus postponed in the proof section.

Proposition 3.4.10 (Upper bound for the minimax risk)**:**
Let $s, L \in \mathbb{R}_+$ and $c \in \mathbb{R}$. Let additionally $\mathbb{E}(U_1^{2(c-1)}) < \infty$ and **[G1]** be fulfilled. Further let $r > 0 \vee (\gamma^{-1} - 1)$. Then for the choice $\eta_{or} := \gamma/(2s + 2\gamma + 1)$, $\rho_n = n^{-\gamma/(2\gamma+2s+1)}$,

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_{f_Y}^n(\|\widehat{f}_{\rho_n,r} - f\|_{x^{2c-1}}^2) \leq C(L, s, g, r) n^{-2s/(2s+2\gamma+1)},$$

where $C(L, s, g, r)$ is a positive constant depending on L, s, r and g.

Proposition 3.4.10 combined with Theorem 3.2.7 implies that, if g fulfills both **[G1]** and **[G1']** the ridge estimator with the choice $\rho_n = n^{-\gamma/2(2\gamma+2s+1)}$ leads to an minimax optimal estimator. Similar to the result of Section 3.2.3, this estimator is again dependent on the unknown regularity parameter $s \in \mathbb{R}_+$. Thus, a fully data-driven choice of η, respectively ρ_n is is needed which we proposed in the next section.

Remark 3.4.11 (Oscillating error densities)**:**
For the case of oscillating error densities, Remark 3.4.8 suggests to choose a value of $c \in \mathbb{R}$ such that the error densities is not oscillating, if possible. In this situation, and if $\kappa = 0$, the above presented result for smooth error densities can be used.
Nevertheless, Remark 3.4.9 indicates, that there are situations where a choice of $c = 1$ is unavoidable resulting in an oscillating error density. The case of $c = 1$ can then be covered by exploiting the theory of oscillating error densities for the Fourier deconvolution in combination with a log transformation of the data, as frequently discussed. Here, we refer to the work of Hall and Meister (2007) for further reading.
For the counterexample proposed in Remark 3.4.9, the expected minimax rate is then given by

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_{f_Y}^n(\|\widehat{f}_{\rho_n} - f\|_x^2) \leq C(L, s, g, r)(\log(n))^{-2\beta/\gamma}$$

for $\eta_{or} = 1/8$ and $r = 0$.

3.4.4 Data-driven choice

We will now present a data-driven version of the estimator $\widehat{f}_{\rho_n,r}$. To do so, let us start by rewriting the estimator $\widehat{f}_{\rho_n,r}$ in a more simple form. We define the family of ridge estimators $(\widehat{f}_{\ell,r})_{\ell \in \mathbb{R}_+}$

through

$$\widehat{f}_{\ell,r}(x) := \frac{1}{2\pi}\int_{\mathbb{R}} x^{-c-it}\widehat{\mathcal{M}}_c(t)\mathrm{R}_{\ell,r,g}(t)d\lambda(t), \quad \mathrm{R}_{\ell,r,g} := \frac{\overline{\mathcal{M}_c[g]}|\mathcal{M}_c[g]|^r}{(|\mathcal{M}_c[g]| \vee \ell^{-1})^{r+2}}, \tag{3.26}$$

where we choose $r > 0$ such that the function $\mathrm{R}_{\ell,r,g} \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. For the data-driven choice of $\ell \in \mathbb{R}_+$ we will use a Goldenshlugger Lepski method. In fact, we will define

$$\widehat{\ell} := \arg\min_{\ell \in \mathcal{L}_n} \widehat{A}(\ell) + \chi_2 \widehat{V}(\ell) \tag{3.27}$$

for two random functions $\widehat{A}, \widehat{V} : \mathcal{L}_n \to \mathbb{R}_+$, a subset $\mathcal{L}_n \subseteq \mathbb{R}_+$ which is rich enough and a $\chi_1 > 0$. Generally spoken, the random function $\widehat{A}$ should behave like the unknown bias of the estimator while the term $\widehat{V}$ should mimic the variance term. Based on Proposition 3.4.3 we define the variance term $V(\ell)$ for $\ell \in \mathbb{R}_+$ by

$$V(\ell) := \frac{\sigma_c}{n}\widetilde{\Delta}_{g,r}(\ell) := \frac{\sigma_c}{n}\|\mathrm{R}_{\ell,r}\|^2_{\mathbb{R}}$$

and its empirical counterpart by $\widehat{V}(\ell) := \frac{2\widehat{\sigma}_c}{n}\widetilde{\Delta}_{g,r}(\ell)$ for $\ell \in \mathbb{R}_+$ and $\widehat{\sigma}_c := n^{-1}\sum_{j=1}^n Y_j^{2(c-1)}$. Considering the variance term, it seems natural to set $\mathcal{L}_n := \{\ell \in \mathbb{N} : \widetilde{\Delta}_{g,r}(\ell) \leq n\}$ and define $L_n := \max(\mathcal{L}_n)$. Further let us define

$$A(\ell) := \sup_{\ell' \in \mathcal{L}_n} (\|\widehat{f}_{\ell',r} - \widehat{f}_{\ell' \wedge \ell, r}\|^2_{x^{2c-1}} - \chi_1 V(\ell'))_+$$

for $\chi_2 \in \mathbb{R}_+$ and $\widehat{A}(\ell) := \sup_{\ell' \in \mathcal{L}_n}(\|\widehat{f}_{\ell',r} - \widehat{f}_{\ell' \wedge \ell, r}\|^2_{x^{2c-1}} - \chi_1 \widehat{V}(\ell'))_+$. Then we can show the following result.

Theorem 3.4.12 (Data-driven ridge estimator)**:**
Let $c \in \mathbb{R}$ and $f \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$. Assume that $\mathbb{E}_{f_Y}(Y_1^{5(c-1)}) < \infty$, $\|gx^{2c-1}\|_\infty < \infty$ and **[G1]** is fulfilled. Then for $\chi_2 \geq \chi_1 \geq 72$ holds

$$\mathbb{E}^n_{f_Y}(\|\widehat{f}_{\widehat{\ell},r} - f\|^2_{x^{2c-1}}) \leq C_1 \inf_{\ell \in \mathcal{L}_n} \left(\|\mathbb{1}_{G_\ell}\mathcal{M}_c[f]\|^2_{\mathbb{R}} + V(\ell)\right) + \frac{C_2}{n},$$

where C_1 is a positive constant depending on χ_2, χ_1 and C_2 is a positive constant depending on $\mathbb{E}_{f_Y}(Y_1^{5(c-1)})$, $\|x^{2c-1}g\|_\infty$, g and r.

Proof of Theorem 3.4.12. The proof can be split in two main steps. The first one is using a sequence of elementary steps to find a controllable upper bound for the risk of the data-driven estimator. In the second step, we use mainly the Talagrand inequality to show the claim of the theorem. These two steps are expressed through the following Lemmata which we state here and postpone the proofs in the appendix.

Lemma 3.4.13:
Under the assumptions of Theorem 3.4.12 holds for any $\ell \in \mathcal{L}_n$

$$\begin{aligned}\mathbb{E}^n_{f_Y}(\|\widehat{f}_{\widehat{\ell},r} - f\|^2_{x^{2c-1}}) \leq& C(\chi_1,\chi_2)\left(\|f - f_{\ell,r}\|^2_{x^{2c-1}} + V(\ell)\right) + C(\chi_1)\mathbb{E}^n_{f_Y}(\sup_{\ell \in \mathcal{L}_n}(\widehat{V}(\ell) - V(\ell))_+) \\ &+ 108\mathbb{E}^n_{f_Y}(\sup_{\ell' \in \mathcal{L}_n}\left(\|\widehat{f}_{\ell',r} - f_{\ell',r}\|^2_{x^{2c-1}} - \frac{\chi_1}{6}V(\ell')\right)_+)\end{aligned}$$

for positive constants $C(\chi_1,\chi_2)$ and $C(\chi_1)$ only depending on χ_1 and χ_2 and $f_{\ell,r} := \mathbb{E}^n_{f_Y}(\widehat{f}_{\ell,r})$.

To be able to apply the Talagrand inequality on the third summand we need to split the process first. To do so, let us define the set $\mathbb{U} := \{h \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) : \|h\|_{x^{2c-1}} \leq 1\}$. Then for $\ell \in \mathcal{L}_n$ we have $\|\widehat{f}_{\ell,r} - f_{\ell,r}\|_{x^{2c-1}} = \sup_{h\in\mathbb{U}} \langle \widehat{f}_{\ell,r} - f_{\ell,r}, h\rangle_{x^{2c-1}}$, where

$$\langle \widehat{f}_{\ell,r} - f_{\ell,r}, h\rangle_{x^{2c-1}} = \frac{1}{2\pi}\int_{\mathbb{R}} \left(\widehat{\mathcal{M}}_c(t) - \mathbb{E}^n_{f_Y}(\widehat{\mathcal{M}}_c(t))\right) \mathrm{R}_{\ell,r,g}(t)\mathcal{M}_c[h](-t)d\lambda(t)$$

by application of the Plancherel identity, Proposition 2.3.5. Now for a positive sequence $(d_n)_{n\in\mathbb{N}}$ we decompose the estimator $\widehat{\mathcal{M}}_c(t)$ into

$$\begin{aligned}\widehat{\mathcal{M}}_c(t) :&= n^{-1}\sum_{j=1}^n Y_j^{c-1+it}\mathbb{1}_{(0,d_n)}(Y_j^{c-1}) + n^{-1}\sum_{j=1}^n Y_j^{c-1+it}\mathbb{1}_{[d_n,\infty)}(Y_j^{c-1})\\ &=: \widehat{\mathcal{M}}_{c,1}(t) + \widehat{\mathcal{M}}_{c,2}(t).\end{aligned}$$

Setting

$$\nu_{\ell,i}(h) := \frac{1}{2\pi}\int_{\mathbb{R}} \left(\widehat{\mathcal{M}}_{c,i}(t) - \mathbb{E}(\widehat{\mathcal{M}}_{c,i}(t))\right) \mathrm{R}_{\ell,r,g}(t)\mathcal{M}_c[h](-t)d\lambda(t) \quad \text{for } h \in \mathbb{U},\ i \in \{1,2\}$$

we can deduce that

$$\begin{aligned}\mathbb{E}^n_{f_Y}(\sup_{\ell\in\mathcal{L}_n}\left(\|\widehat{f}_{\ell,r} - f_{\ell,r}\|^2_{x^{2c-1}} - \frac{\chi_1}{6}V(\ell)\right)_+) &\leq 2\mathbb{E}^n_{f_Y}(\sup_{\ell\in\mathcal{L}_n}\left(\sup_{h\in\mathbb{U}}\nu_{\ell,1}(h)^2 - \frac{\chi_1}{12}V(\ell)\right)_+)\\ &\quad + 2\mathbb{E}^n_{f_Y}(\sup_{\ell\in\mathcal{L}_n}\sup_{h\in\mathbb{U}}\nu_{\ell,2}(h)^2)).\end{aligned} \tag{3.28}$$

This decomposition and the following Lemma proves then the claim.

Lemma 3.4.14:
Under the assumptions of Theorem 3.4.12 holds

$$(i)\quad \mathbb{E}^n_{f_Y}(\sup_{\ell\in\mathcal{L}_n}(\sup_{h\in\mathbb{U}}\nu_{\ell,1}(h)^2 - \frac{\chi_1}{12}V(\ell))_+) \leq \frac{C(g, r, \mathbb{E}_f(X_1^{2(c-1)})}{n},$$

$$(ii)\quad \mathbb{E}^n_{f_Y}(\sup_{\ell\in\mathcal{L}_n}\sup_{h\in\mathbb{U}}\nu_{\ell,2}(h)^2)) \leq \frac{C(\sigma_c, \mathbb{E}_{f_Y}(Y_1^{5(c-1)}))}{n},$$

$$(iii)\quad \mathbb{E}^n_{f_Y}(\sup_{\ell\in\mathcal{L}_n}(\widehat{V}(\ell) - V(\ell))_+) \leq \frac{C(\mathbb{E}_{f_Y}(Y_1^{4(c-1)}), \mathbb{E}_{f_Y}(Y_1^{2(c-1)}))}{n}.$$

□

Assuming again that the density lies in a Mellin-Sobolev ellipsoid, we can deduce directly the following Corollary whose proof is thus omitted.

Corollary 3.4.15:
Let $c \in \mathbb{R}$, $s, L \in \mathbb{R}_+$ and $f \in \mathbb{D}^s_c(L)$. Assume further that $\mathbb{E}(Y_1^{5(c-1)}) < \infty$, $\|gx^{2c-1}\|_\infty < \infty$ and **[G1]** is fulfilled. Then for $\chi_2 \geq \chi_1 \geq 72$ holds

$$\mathbb{E}^n_{f_Y}(\|\widehat{f}_{\widehat{\ell},r} - f\|^2_{x^{2c-1}}) \leq C(L, s, r, g, \mathbb{E}(X_1^{5(c-1)}))\, n^{-2s/(2s+2\gamma+1)},$$

where $C(L, s, r, g, \mathbb{E}(X_1^{5(c-1)}))$ is a positive constant depending on L, s, r, g and $\mathbb{E}(X_1^{5(c-1)})$.

3.4.5 Numerical results

Since both, the data-driven ridge estimator $\widehat{f}_{\widehat{\ell},r}$ and the data-driven spectral cut-off estimator $\widehat{f}_{\widehat{k}}$ are densities estimators, we will illustrate the behavior of the data-driven ridge estimator $\widehat{f}_{\widehat{\ell},r}$, presented in (3.26) and (3.27), in comparison to the data-driven spectral cut-off estimator $\widehat{f}_{\widehat{k}}$, presented in equation (3.14) and (3.15). To do so, we use the following examples for the unknown density f, which has been already considered in Section 3.2.5.

(i) Beta Distribution: $f(x) = B(2,5)^{-1}x(1-x)^4\mathbb{1}_{(0,1)}(x), x \in \mathbb{R}_+$,

(ii) Log-Gamma Distribution: $f(x) = 5^5\Gamma(5)^{-1}x^{-6}\log(x)^4\mathbb{1}_{(1,\infty)}(x), x \in \mathbb{R}_+$,

(iii) Gamma Distribution: $f(x) = \Gamma(5)^{-1}x^4\exp(-x)\mathbb{1}_{(0,\infty)}(x), x \in \mathbb{R}_+$ and

(iv) Log-Normal Distribution: $f(x) = (0.32\pi x^2)^{-1/2}\exp(-\log(x)^2/0.32)\mathbb{1}_{(0,\infty)}(x), x \in \mathbb{R}_+$.

A detailed discussion of these examples in terms of the decay of their Mellin transform and its decay can be found in section 2.5. To visualize the behavior of the estimator, we use the following examples of error densities g, namely

$a)$ Symmetric noise: $g(x) = \mathbb{1}_{(0.5,1.5)}(x), x \in \mathbb{R}_+$,

$b)$ Beta Distribution: $g(x) = 2x\mathbb{1}_{(0,1)}(x), x \in \mathbb{R}_+$,

$c)$ Uniform Distribution: $g(x) = \mathbb{1}_{(0,1)}(x), x \in \mathbb{R}_+$ and

$d)$ Pareto Distribution: $g(x) = \mathbb{1}_{(1,\infty)}(x)x^{-2}$, $x \in \mathbb{R}_+$.

Here it is worth pointing out that the example $a), c)$ and $d)$ fulfill **[G1]** with $\gamma = 1$ and $b)$ with $\gamma = 2$. By minimizing an integrated weighted squared error over a family of histogram densities with randomly drawn partitions and weights, we select $\gamma = 1$ $\chi_1 = \chi_2 = 72$ for $\widehat{f}_{\widehat{\ell},r}$ and $\chi = 5$ for $\widehat{f}_{\widehat{k}}$. For the case $b)$ we choose $\chi_1 = \chi_2 = 6$ and $\chi = 3$. In both cases, we have set $r = 2$.
Figure 3.10 shows that both estimators behave similar. As suggested by the theory, the reconstruction of the density f from the observation $Y_1, \dots, Y_n$ seems to be more difficult, if the error variable is uniformly distributed, case $a)$, than if the error variable is Beta distributed, case $b)$.
Again we see that both estimator react analogously to varying values of the model parameter $c \in \mathbb{R}$. Looking at the medians in Figure 3.11, for $c = 0$ the median seems to be closer to the true density for smaller values of $x \in \mathbb{R}_+$. For $c = 1$ the opposite effects seems to occur. For $c = 1/2$, the case of the unweighted $\mathbb{L}^2$-distance, such effects cannot be observed. Regarding the risk, this seems natural as the weight function for $c = 0$ is monotonically decreasing, while for $c = 1$ it is monotonically increasing.

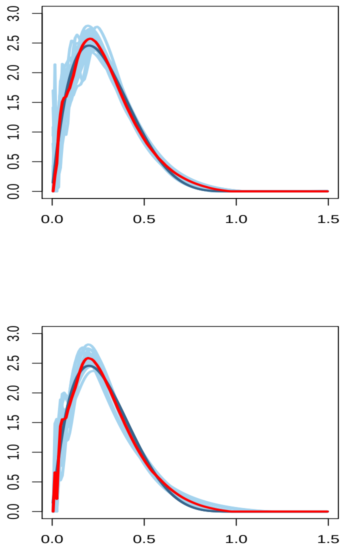

Figure 3.10: The estimator $\widehat{f}_{\widehat{\ell},r}$ (top) and $\widehat{f}_{\widehat{k}}$ (bottom) is depicted for 50 Monte-Carlo simulations with sample size $n = 2000$ in the case (i) under the error density $a)$ (left) and $b)$ (right) for $c = 1$. The true density f is given by the black curve while the red curve is the point-wise empirical median of the 50 estimates.

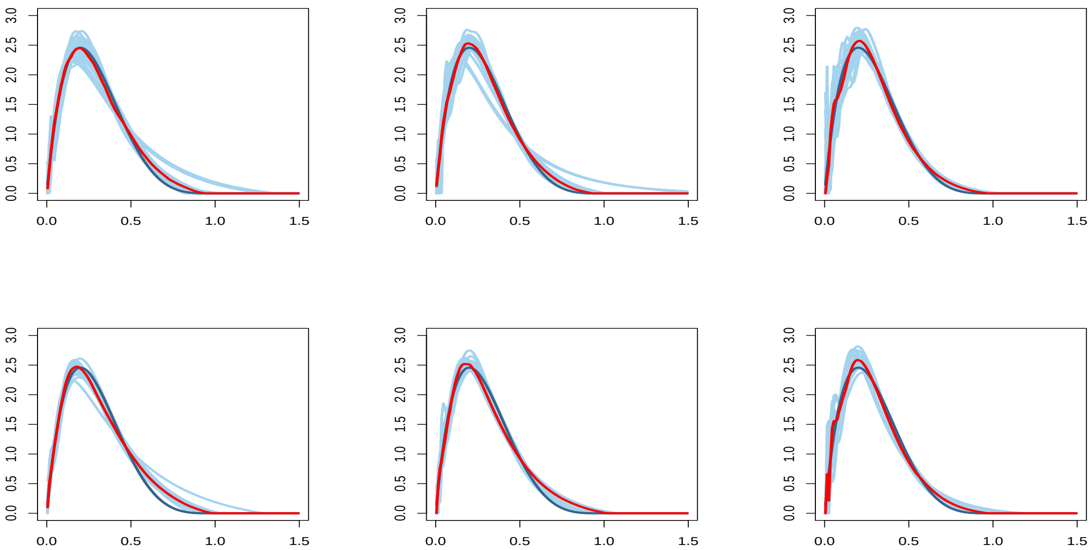

Figure 3.11: The estimator $\widehat{f}_{\widehat{\ell},r}$ (top) and $\widehat{f}_{\widetilde{k}}$ (bottom) is depicted for 50 Monte-Carlo simulations with sample size $n = 2000$ in the case (i) under the error density $a)$ for $c = 0$ (left), $c = 1/2$ (middle) and $c = 1$ (right). The true density f is given by the black curve while the red curve is the point-wise empirical median of the 50 estimates.

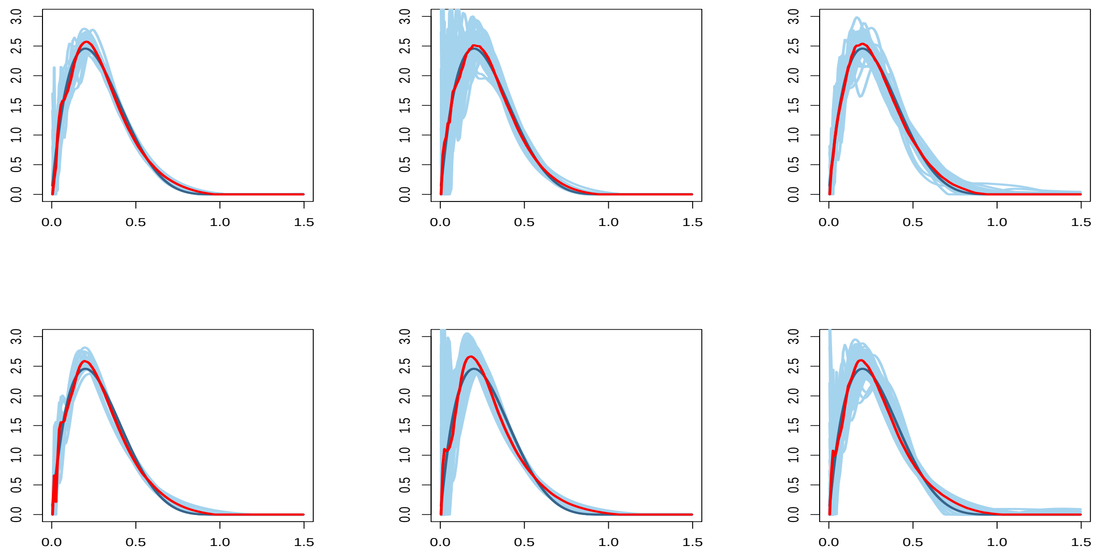

Figure 3.12: The estimator $\widehat{f}_{\widehat{\ell},r}$ (top) and $\widehat{f}_{\widetilde{k}}$ (bottom) is depicted for 50 Monte-Carlo simulations with sample size $n = 2000$ in the case (i) under the error density $a)$ (left), $c)$ (middle) and $d)$ (right). The true density f is given by the black curve while the red curve is the point-wise empirical median of the 50 estimates.

In Figure 3.12 we see that both estimator react similar for different examples of error densities with common error parameter $\gamma \in \mathbb{R}_+$, $\gamma = 1$.

Based on the results of the following Table 3.1, none of the estimators performance strictly better in terms of MISE. Nevertheless, we see again:

	Case	(i)		(ii)		(iii)		(iv)	
	Sample size	500	2000	500	2000	500	2000	500	2000
$a)$	Ridge	0.94	0.31	2.17	1.54	0.63	0.17	7.13	2.38
	Spectral	1.10	0.38	2.03	1.26	0.52	0.16	15.07	2.34
$b)$	Ridge	2.32	1.43	5.90	3.81	1.19	0.47	25.84	11.03
	Spectral	3.95	1.56	10.63	7.12	1.52	0.84	33.95	13.45

Table 3.1: The entries showcase the MISE (scaled by a factor of 100) obtained by Monte-Carlo simulations each with 500 iterations. We take a look at different densities f and g, two distinct sample sizes, and for both estimators $\widehat{f}_{\widehat{k}}$ and $\widetilde{f}_{\widetilde{k}}$ for $c = 1$.

3.4.6 Conclusion

In this section we have considered the estimation of the density f given an i.i.d. sample of $Y_1, \ldots, Y_n$. In contrary to Section 3.2, we constructed an estimator using a Ridge approach and showed its consistency under the minimal assumption **[G-1]**. We proposed examples of error densities fulfilling **[G-1]**, but not **[G0]**, the so-called oscillating error densities. Afterwards we focused througout the section on the case of smooth error densities and showed that our estimation procedure is minimax-optimal and proposed a data-driven choice of the ridge parameter. We finished the section by illustrating the finite sample properties of our estimator using a Monte-Carlo simulation in comparison to the data-driven spectral cut-off estimator.

3.4.7 Proofs

Proof Proposition 3.4.3. We start by considering the bias term. In fact, we have for $t \in G_n^c$ $\mathrm{R}_{\rho_n,r,g}(t) = \mathcal{M}_c[g](t)^{-1}$. On the other hand, for $t \in G_n$ we have

$$|\mathrm{R}_{\rho_n,r,g}(t)| = \frac{|\mathcal{M}_c[g](t)|^{r+1}}{(|\mathcal{M}_c[g]| \vee \rho_n)(t))^{r+2}} \leq |\mathcal{M}_c[g](t)|^{-1}.$$

Now using the Plancherel identity, compare Proposition 2.3.5,

$$\begin{aligned}
\|f - \mathbb{E}(\widehat{f}_{\rho_n,r})\|_{x^{2c-1}}^2 &= \frac{1}{2\pi}\int_{\mathbb{R}} |\mathcal{M}_c[g](t)\mathrm{R}_{\rho_n,r,g}(t) - 1|^2 \, |\mathcal{M}_c[f](t)|^2 d\lambda(t) \\
&= \frac{1}{2\pi}\int_{G_n} |\mathcal{M}_c[g](t)\mathrm{R}_{\rho_n,r,g}(t) - 1|^2 \, |\mathcal{M}_c[f](t)|^2 d\lambda(t) \\
&\leq \int_{G_n} |\mathcal{M}_c[f](t)|^2 d\lambda(t).
\end{aligned}$$

Let us now consider the variance term. In fact, we see directly that

$$\begin{aligned}
\frac{1}{2\pi}\int_{\mathbb{R}} \mathbb{V}\mathrm{ar}_{f_Y}^n(\widehat{\mathcal{M}}_c(t))|\mathrm{R}_{\rho_n,r,g}(t)|^2 d\lambda(t) &= \frac{1}{2\pi n}\int_{\mathbb{R}} \mathbb{V}\mathrm{ar}_{f_Y}(Y_1^{c-1+it})|\mathrm{R}_{\rho_n,r,g}(t)|^2 d\lambda(t) \\
&\leq \frac{\sigma_c}{2\pi n}\|\mathrm{R}_{\rho_n,r,g}\|_{\mathbb{R}}^2.
\end{aligned}$$

□

Proof of Proposition 3.4.10. For the ridge estimator, using a simple computation we see that if g fulfills **[G1]**, then we can find positive constants $C_{g,1}, C_{g,2} > 0$ only depending on g such that the sets $G_{n,i} := \mathbb{R} \setminus [-C_{g,i}n^{\eta/\gamma}, C_{g,i}n^{\eta/\gamma}]$ for $i = 1, 2$ fulfill the inclusion relationship

$$G_{n,1} \subseteq G_n \subseteq G_{n,2}.$$

This relationship will help us to control the summands of the bound given in Proposition 3.4.3

$$\begin{aligned}
\mathbb{E}_{f_Y}^n(\|f - \widehat{f}_{\rho_n,r}\|_{x^{2c-1}}^2) &\leq \|\mathbb{1}_{G_n}\mathcal{M}_c[f]\|_{\mathbb{R}}^2 + \frac{\sigma_c}{n}\int_{\mathbb{R}} \frac{\mathbb{1}_{G_n}(t)|\mathcal{M}_c[g](t)|^{2r+2}}{\rho_n^{2r+4}} + \frac{\mathbb{1}_{G_n^c}(t)}{|\mathcal{M}_c[g](t)|^2} d\lambda(t) \\
&\leq \|\mathbb{1}_{G_{n,2}}\mathcal{M}_c[f]\|_{\mathbb{R}}^2 + \frac{\sigma_c}{n}\left\|\frac{\mathbb{1}_{G_{n,2}}\mathcal{M}_c[g]^{r+1}}{\rho_n^{r+2}}\right\|_{\mathbb{R}}^2 + \frac{\sigma_c}{n}\left\|\frac{\mathbb{1}_{G_{n,1}^c}}{\mathcal{M}_c[g]}\right\|_{\mathbb{R}}^2.
\end{aligned}$$

As $f \in \mathbb{D}_c^s(L)$ we get

$$\|\mathbb{1}_{G_{n,2}}\mathcal{M}_c[f]\|_{\mathbb{R}}^2 = \int_{(C_{g,2}n^{\eta/\gamma},\infty)} |\mathcal{M}_c[f](t)|^2 d\lambda(t) \leq C(g,L)n^{-2s\eta/\gamma}.$$

For the second summand we see, since $\gamma(r+1) > 1$,

$$\begin{aligned}\frac{\sigma_c}{n}\left\|\frac{\mathbb{1}_{G_{n,2}}\mathcal{M}_c[g]^{r+1}}{\rho_n^{r+2}}\right\|_{\mathbb{R}}^2 &= C(g,L)n^{2\eta(r+2)-1}\int_{(C_{g,2}n^{\eta/\gamma},\infty)} t^{-2\gamma(r+1)}dt\\ &= C(g,L,r)n^{2\eta(r+2)-1-2\eta(r+1)+\eta/\gamma}\\ &= C(g,L,r)n^{2\eta+\eta/\gamma-1}.\end{aligned}$$

For the third summand we get

$$\frac{\sigma_c}{n}\int_{(-C_{g,1}n^{\eta/\gamma},C_{g,1}n^{\eta/\gamma}}|\mathcal{M}_c[g](t)|^2 d\lambda(t) \leq C(g,L)n^{2\eta+\eta/\gamma-1}.$$

In total we have

$$\mathbb{E}(\|f-\widehat{f}_{\rho_n,r}\|_{x^{2c-1}}^2) \leq C(g,L,r)(n^{-2s\eta/\gamma}+n^{2\eta+\eta/\gamma-1}),$$

where both summands are balanced for the choice $\eta = \gamma/(2s+2\gamma+1)$. □

Proof of Lemma 3.4.13. In fact, by the definition of $\widehat{\ell}$ it follows for any $\ell \in \mathcal{L}_n$ and since $\chi_2 \geq \chi_1$,

$$\begin{aligned}\|f-\widehat{f}_{\widehat{\ell},r}\|_{x^{2c-1}}^2 &\leq 3\|f-\widehat{f}_{\ell,r}\|_{x^{2c-1}}^2 + 3\|\widehat{f}_{\ell,r}-\widehat{f}_{\ell\wedge\widehat{\ell},r}\|_{x^{2c-1}}^2 + 3\|\widehat{f}_{\ell\wedge\widehat{\ell},r}-\widehat{f}_{\widehat{\ell},r}\|_{x^{2c-1}}^2\\ &\leq 3\|f-\widehat{f}_{\ell,r}\|_{x^{2c-1}}^2 + 3(\widehat{A}(\widehat{\ell})+\chi_1\widehat{V}(\ell)+\widehat{A}(\ell)+\chi_1\widehat{V}(\widehat{\ell}))\\ &\leq 3\|f-\widehat{f}_{\ell,r}\|_{x^{2c-1}}^2 + 3(2\widehat{A}(\ell)+(\chi_1+\chi_2)\widehat{V}(\ell)).\end{aligned}$$

To simplify the notation, we set $\chi := (\chi_1+\chi_2)/2$. Let us now have a closer look at $\widehat{A}(\ell)$. From

$$\begin{aligned}\|\widehat{f}_{\ell',r}-\widehat{f}_{\ell'\wedge\ell,r}\|_{x^{2c-1}}^2 &\leq 3(\|\widehat{f}_{\ell',r}-f_{\ell',r}\|_{x^{2c-1}}^2 + \|\widehat{f}_{\ell'\wedge\ell,r}-f_{\ell'\wedge\ell,r}\|_{x^{2c-1}}^2 + \|f_{\ell,r}-f_{\ell'\wedge\ell,r}\|_{x^{2c-1}}^2)\\ &\leq 6\|\widehat{f}_{\ell',r}-f_{\ell',r}\|_{x^{2c-1}}^2 + 3\|f-f_{\ell,r}\|_{x^{2c-1}}^2\end{aligned}$$

we conclude by a straight forward computation and the monotony of V that

$$\widehat{A}(\ell) \leq 6\sup_{\ell'\in\mathcal{L}_n}\left(\|\widehat{f}_{\ell',r}-f_{\ell',r}\|_{x^{2c-1}}^2 - \frac{\chi_1}{6}V(\ell')\right)_+ + 3\|f-f_{\ell,r}\|_{x^{2c-1}}^2 + \chi_1\sup_{\ell'\in\mathcal{L}_n}(\widehat{V}(\ell')-V(\ell'))_+.$$

This implies

$$\begin{aligned}\mathbb{E}_{f_Y}^n(\|f-\widehat{f}_{\widehat{\ell},r}\|_{x^{2c-1}}^2) \leq C(\chi)\left(\|f-f_{\ell,r}\|_{x^{2c-1}}^2+V(\ell)\right) + C(\chi_1)\mathbb{E}_{f_Y}^n(\sup_{\ell\in\mathcal{L}_n}(\widehat{V}(\ell)-V(\ell))_+)\\ + 108\mathbb{E}_{f_Y}^n(\sup_{\ell'\in\mathcal{L}_n}(\|\widehat{f}_{\ell',r}-f_{\ell',r}\|_{x^{2c-1}}^2 - \frac{\chi_1}{6}V(\ell'))_+).\end{aligned}$$

□

Proof of Lemma 3.4.14. Proof of (i): Now on the first summand of the right hand side of (3.28) we can apply the Talagrand. We use that

$$\mathbb{E}_{f_Y}^n(\sup_{\ell\in\mathcal{L}_n}(\sup_{h\in\mathbb{U}}\nu_{\ell,1}(h)^2 - \frac{\chi_1}{12}V(\ell))_+) \leq \sum_{\ell=1}^{L_n}\mathbb{E}_{f_Y}^n((\sup_{h\in\mathbb{U}}\nu_{\ell,1}(h)^2 - \frac{\chi_1}{12}V(\ell))_+).$$

To apply the Talagrand inequality, compare Lemma 3.1.14, to each summand we need to determine the constants Ψ^2, ψ^2 and τ first. Staying in the notation of the Talagrand inequality, we set for $h \in \mathbb{U}$,

$$\nu_h(y) := \frac{1}{2\pi} \int_{\mathbb{R}} y^{c-1+it} \mathbb{1}_{(0,d_n)}(y^{c-1}) \mathrm{R}_{\ell,r,g}(t) \mathcal{M}_c[h](-t) d\lambda(t)$$

for any $y > 0$. Now applying Cauchy-Schwarz inequality $\nu_{\ell,1}(h)^2 \leq \|\widehat{f}_{\ell,r} - f_{\ell,r}\|^2_{x^{2c-1}} \|h\|^2_{x^{2c-1}} \leq \|\widehat{f}_{\ell,r} - f_{\ell,r}\|^2_{x^{2c-1}}$ since $h \in \mathbb{U}$. We then deduce

$$\mathbb{E}^n_{f_Y}(\sup_{h \in \mathbb{U}} \nu_{\ell,1}(h)^2) \leq \mathbb{E}^n_{f_Y}(\|\widehat{f}_{\ell,r} - f_{\ell,r}\|^2_{x^{2c-1}}) \leq \sigma_c \widetilde{\Delta}_{g,r}(\ell) n^{-1} =: \Psi^2,$$

compare proof of Proposition 3.4.3. For $y \in \mathbb{R}_+$ we have $|\nu_h(y)|^2 \leq d_n^2 \|\mathrm{R}_{\ell,r,g}\|^2_{\mathbb{R}} \|\mathcal{M}_c[h]\|^2_{\mathbb{R}}/(2\pi)^2 \leq d_n^2 \|\mathrm{R}_{\ell,r,g}\|^2_{\mathbb{R}} \leq d_n^2 \widetilde{\Delta}_{g,r}(\ell) =: \psi^2$ since $h \in \mathbb{U}$. Additionally, for $h \in \mathbb{U}$ holds $\mathbb{V}\mathrm{ar}^n_{f_Y}(\nu_h(Y_1)) \leq \mathbb{E}^n_{f_Y}(\nu_h^2(Y_1)) \leq \|f_Y x^{2c-1}\|_\infty \|\nu_h\|^2_{x^{1-2c}}$. More precisely, we see that

$$y^{2c-1} f_Y(y) = y^{2c-1} \int_{\mathbb{R}_+} f(x) g(y/x) x^{-1} d\lambda(x) \leq \|x^{2c-1} g\|_\infty \mathbb{E}(X_1^{2(c-1)}), \quad \text{for any } y \in \mathbb{R}_+.$$

Next, we have

$$\|\nu_h\|^2_{x^{1-2c}} \leq \frac{1}{2\pi} \int_{\mathbb{R}} |\mathcal{M}_c[h](t)|^2 |\mathrm{R}_{\ell,r,g}(t)|^2 d\lambda(t) \leq \|\mathrm{R}^2_{\ell,r,g}\|_\infty \frac{1}{2\pi} \|\mathcal{M}_c[h]\|^2_{\mathbb{R}} \leq \|\mathrm{R}^2_{\ell,r,g}\|_\infty,$$

which implies the choice $\tau := \|x^{2c-1} g\|_\infty \mathbb{E}_f(X_1^{2(c-1)}) \|\mathrm{R}^2_{\ell,r,g}\|_\infty$. Applying the Talagrand inequality we get

$$\begin{aligned}
\mathbb{E}^n_{f_Y}((\sup_{h \in \mathbb{U}} \overline{\nu}_h^2 - 6\Psi^2)_+) &\leq \frac{C_{f_Y}}{n} \left(\|\mathrm{R}_{\ell,r,g}\|_\infty \exp(-C_{f_Y} \frac{\widetilde{\Delta}_{g,r}(\ell)}{\|\mathrm{R}_{\ell,r,g}\|_\infty}) + d_n^2 \exp(-\frac{\sqrt{n\sigma_c}}{100 d_n}) \right) \\
&\leq \frac{C_{f_Y}}{n} \left(\|\mathrm{R}_{\ell,r,g}\|_\infty \exp(-C_{f_Y} \frac{\widetilde{\Delta}_{g,r}(\ell)}{\|\mathrm{R}_{\ell,r,g}\|_\infty}) + n^{-1} \right)
\end{aligned}$$

for the choice $d_n := \sqrt{n\sigma}/(100 \log(n^2))$. Following the same step as in the proof of Proposition 3.4.10, we can state that $L_n \leq C_{g,r} n^{\gamma/(2\gamma+1)} \leq C_{g,r} n^1$. For $\chi_1 \geq 72$ we can conclude that

$$\begin{aligned}
\mathbb{E}^n_{f_Y}(\sup_{\ell \in \mathcal{L}_n} \left(\sup_{h \in \mathbb{U}} \nu_{\ell,1}(h)^2 - \frac{\chi_1}{12} V(\ell) \right)_+) &\leq \sum_{\ell=1}^{L_n} \frac{C_{f_Y}}{n} \left(\|\mathrm{R}_{\ell,r,g}\|_\infty \exp(-C_{f_Y} \frac{\widetilde{\Delta}_{g,r}(\ell)}{\|\mathrm{R}_{\ell,r,g}\|_\infty}) + n^{-1} \right) \\
&\leq \frac{C_{f_Y}}{n} (1 + \sum_{k=1}^{K_n} \|\mathrm{R}_{\ell,r,g}\|_\infty \exp(-C_{f_Y} \frac{\widetilde{\Delta}_{g,r}(\ell)}{\|\mathrm{R}_{\ell,r,g}\|_\infty})).
\end{aligned}$$

Now it can easily seen that there exist constants $c_{g,r}, C_{g,r} > 0$ such that $c_{g,r} \ell^{2+\gamma^{-1}} \leq \widetilde{\Delta}_{g,r}(k) \leq C_g \ell^{2+\gamma^{-1}}$ using **[G1]**. By a simple computation, we can show that $\|\mathrm{R}_{\ell,r,g}\|_\infty \leq C_{g,r} \ell^2$. Now, since $(\ell^2 e^{-C_{f_Y} \ell^{1/\gamma}})_{\ell \in \mathbb{N}}$ is summable we get $\mathbb{E}^n_{f_Y}(\sup_{\ell \in \mathcal{L}_n} (\sup_{h \in \mathbb{U}} \nu_{\ell,1}(h)^2 - \frac{\chi_1}{12} V(\ell))_+) \leq C_{f_Y} n^{-1}$.
Let us now show part *(ii)* : For any $h \in \mathbb{U}$ and $\ell \in \mathcal{L}_n$ we get

$$\nu_{\ell,2}(h)^2 \leq (2\pi)^{-1} \|\widehat{\mathcal{M}}_{c,2} \mathrm{R}_{\ell,r,g}\|^2_{\mathbb{R}} \|h\|^2_{x^{2c-1}} \leq \|\widehat{\mathcal{M}}_{c,2} \mathrm{R}_{\ell,r,g}\|^2_{\mathbb{R}}.$$

Therefore

$$\mathbb{E}^n_{f_Y}\big(\sup_{\ell\in\mathcal{L}_n}\sup_{h\in\mathbb{U}}\nu_{\ell,2}(h)^2)\big) \leq \mathbb{E}^n_{f_Y}\big(\sup_{\ell\in\mathcal{L}_n}\|\widehat{\mathcal{M}}_{c,2}\mathrm{R}_{\ell,r,g}\|^2_{\mathbb{R}}\big) \leq \frac{1}{n}\|\mathrm{R}_{L_n,r,g}\|^2_{\mathbb{R}}\mathbb{E}_{f_Y}(Y_1^{2(c-1)}\mathbb{1}_{[d_n,\infty)}(Y_1))$$

since by the definition of $\mathrm{R}_{\ell,r,g}$ it is clear that $\mathrm{R}_{k,r,g}(t) \geq \mathrm{R}_{\ell,r,g}(t)$ for all $t \in \mathbb{R}$ and $k \geq \ell$. Now by definition of $\mathcal{L}_n$ we know that $\|\mathrm{R}_{L_n,r,g}\|^2_{\mathbb{R}}n^{-1} \leq 1$. We deduce that for any $p > 0$ it holds

$$\mathbb{E}^n_{f_Y}\big(\sup_{\ell\in\mathcal{L}_n}\sup_{h\in\mathbb{U}}\nu_{\ell,2}(h)^2)\big) \leq d_n^{-p}\mathbb{E}_{f_Y}(Y_1^{(2+p)(c-1)}) \leq \frac{C(\sigma_c, \mathbb{E}_{f_Y}(Y_1^{5(c-1)}))}{n}$$

choosing $p = 3$ and by the definition of $(c_n)_{n\in\mathbb{N}}$.
Part *(iii)*: First we see that for any $\ell \in \mathcal{L}_n$ it holds $(V(\ell) - \widehat{V}(\ell))_+ = \widetilde{\Delta}_{g,r}(k)n^{-1}(\sigma_c - 2\widehat{\sigma}_c)_+ \leq (\sigma_c - 2\widehat{\sigma}_c)_+$. On $\Omega := \{|\widehat{\sigma}_c - \sigma_c| \leq \sigma_c/2\}$ we have $\frac{\sigma_c}{2} \leq \widehat{\sigma}_c \leq \frac{3}{2}\sigma_c$. This implies

$$\mathbb{E}^n_{f_Y}\big(\sup_{\ell\in\mathcal{L}_n}(V(\ell) - \widehat{V}(\ell))_+\big) \leq \mathbb{E}^n_{f_Y}((\sigma_c - 2\widehat{\sigma}_c)_+) \leq 2\mathbb{E}^n_{f_Y}(|\sigma_c - \widehat{\sigma}_c|\mathbb{1}_{\Omega^c}) \leq 4\frac{\mathbb{V}\mathrm{ar}^n_{f_Y}(\widehat{\sigma}_c)}{\sigma_c}$$

applying the Cauchy-Schwarz inequality and the Markov inequality. Now the last inequality implies the claim. □

3.5 Under multiplicative measurement errors for multivariate densities

3.5.1 Introduction

We will now investigate the multivariate version of the model presented in Section 3.2. More precisely, we are interested in estimating the unknown density $f : \mathbb{R}_+^d \to \mathbb{R}_+$ of a positive random variable $\boldsymbol{X} = (X_1, \dots, X_d)$ given independent and identically distributed (i.i.d.) copies of

$$\boldsymbol{Y} = (X_1U_1, \dots, X_dU_d) =: \boldsymbol{XU},$$

where $\boldsymbol{X}$ and $\boldsymbol{U}$ are independent of each other and $\boldsymbol{U}$ has a known density $g : \mathbb{R}_+^d \to \mathbb{R}_+$. In this setting the density $f_{\boldsymbol{Y}} : \mathbb{R}_+^d \to \mathbb{R}_+$ of $\boldsymbol{Y}$ is given by

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = (f * g)(\boldsymbol{y}) := \int_{\mathbb{R}_+^d} f(\boldsymbol{x}) g(\boldsymbol{y}/\boldsymbol{x}) \boldsymbol{x}^{\underline{-1}} d\lambda^d(\boldsymbol{x}), \quad \text{for } \boldsymbol{y} \in \mathbb{R}_+^d,$$

the multiplicative convolution between the function $f : \mathbb{R}_+^d \to \mathbb{R}_+$ and $g : \mathbb{R}_+^d \to \mathbb{R}_+$, defined in Definition 2.4.1.

In the additive deconvolution literature the density estimation for multivariate variables based on non-parametric estimators has been studied by many authors. A kernel estimator approach was investigated by Comte and Lacour (2013) with respect to $\mathbb{L}^2$-risk and by Rebelles (2016) for general $\mathbb{L}^p$-risk. The multivariate convolution structure density model was considered by the authors of Lepski and Willer (2019). The recent work Dussap (2021) focuses on the study of deconvolution problems on $\mathbb{R}_+^d$ and introduces a data-driven estimator based on a projection on the Laguerre basis. An inclusion of the univariate case in the density estimation literature can be found in section 3.2.1.

The section is organized as follows. In Subsection 3.5.2 we propose the density estimator based on a spectral cut-off approach. Characterizing the regularity of both, the unknown density f and the error density g, by the decay of their corresponding Mellin transform, we develop the minimax optimality of our estimator for smooth error densities in Subsection 3.5.3. A data-driven choice of the upcoming smoothing parameter based on a Goldenshluger-Lepski method is presented and studied in Subsection 3.5.4. Finally, results of a simulation study are reported in Subsection 3.5.5, which visualizes the reasonable finite sample performance of our estimator. The proof of Subsection 3.5.3 and sub section 3.5.4 are postponed to the Subsection 3.5.7.

3.5.2 Estimation strategy

We will now present a generalization of the spectral cut-off approach presented in Section 3.2 for the case of multivariate distributions. Let us recapitulate that for $\boldsymbol{k} \in \mathbb{R}_+^d$ we defined the hyper cuboid $[-\boldsymbol{k}, \boldsymbol{k}] := \{\boldsymbol{x} \in \mathbb{R}^d : \forall i \in [\![d]\!] : |\boldsymbol{x}_i| \le k_i\}$ and for $f \in \mathbb{L}^2(\mathbb{R}_+^d, \boldsymbol{x}^{\underline{2\mathbf{c}-\mathbf{1}}})$ the approximations

$$f_{\boldsymbol{k}}(\boldsymbol{x}) := \frac{1}{(2\pi)^d} \int_{[-\boldsymbol{k},\boldsymbol{k}]} \boldsymbol{x}^{\underline{-\mathbf{c}-i\boldsymbol{t}}} \mathcal{M}_{\boldsymbol{c}}[f](\boldsymbol{t}) d\lambda^d(\boldsymbol{t}), \quad \text{for } \boldsymbol{x} \in \mathbb{R}_+^d.$$

Since $\mathcal{M}_{\boldsymbol{c}}[f]\mathbb{1}_{[-\boldsymbol{k},\boldsymbol{k}]} \in \mathbb{L}^2(\mathbb{R}^d) \cap \mathbb{L}^1(\mathbb{R}^d)$ we get that $f_{\boldsymbol{k}}$ is an approximation of f in the $\mathbb{L}^2(\mathbb{R}_+^d, \boldsymbol{x}^{\underline{2\mathbf{c}-\mathbf{1}}})$-sense, that is $\|f_{\boldsymbol{k}} - f\|_{\boldsymbol{x}^{2\mathbf{c}-\mathbf{1}}} \to 0$ for $\boldsymbol{k} \to \infty$ where the limit $\boldsymbol{k} \to \infty$ means that every component of $\boldsymbol{k}$ is going to infinity, compare (2.4).

Now let us additionally assume that $f \in \mathbb{L}^2(\mathbb{R}_+^d, \boldsymbol{x}^{\underline{2\mathbf{c}-\mathbf{1}}}) \cap \mathbb{L}^1(\mathbb{R}_+^d, \boldsymbol{x}^{\underline{\mathbf{c}-\mathbf{1}}})$ and $g \in \mathbb{L}^1(\mathbb{R}_+^d, \boldsymbol{x}^{\underline{\mathbf{c}-\mathbf{1}}})$.

Then from the convolution theorem, we deduce $\mathcal{M}_{\boldsymbol{c}}[f_{\boldsymbol{Y}}] = \mathcal{M}_{\boldsymbol{c}}[f]\mathcal{M}_{\boldsymbol{c}}[g]$. Under the mild assumption that $\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t}) \neq 0$, for any $\boldsymbol{t} \in \mathbb{R}^d$ we can rewrite the last equation as $\mathcal{M}_{\boldsymbol{c}}[f] = \mathcal{M}_{\boldsymbol{c}}[f_{\boldsymbol{Y}}]/\mathcal{M}_{\boldsymbol{c}}[g]$. Thus we have

$$f_{\boldsymbol{k}}(\boldsymbol{x}) := \frac{1}{(2\pi)^d}\int_{[-\boldsymbol{k},\boldsymbol{k}]} \underline{\boldsymbol{x}}^{-\boldsymbol{c}-i\boldsymbol{t}} \frac{\mathcal{M}_{\boldsymbol{c}}[f_{\boldsymbol{Y}}](\boldsymbol{t})}{\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})} d\lambda^d(\boldsymbol{t}), \qquad \text{for } \boldsymbol{x} \in \mathbb{R}^d_+.$$

Let us now consider for any $\boldsymbol{t} \in \mathbb{R}^d$ the unbiased estimator $\widehat{\mathcal{M}}_c(\boldsymbol{t}) := n^{-1}\sum_{j\in[\![n]\!]} \boldsymbol{Y}_j^{\underline{\boldsymbol{c}-\boldsymbol{1}+i\boldsymbol{t}}}$ of $\mathcal{M}_{\boldsymbol{c}}[f_{\boldsymbol{Y}}](\boldsymbol{t})$. We see that $|\widehat{\mathcal{M}}_c(\boldsymbol{t})| \leq |\widehat{\mathcal{M}}_c(\boldsymbol{0})| < \infty$ almost surely. If additionally $\mathbb{1}_{[-\boldsymbol{k},\boldsymbol{k}]}/\mathcal{M}_{\boldsymbol{c}}[g] \in \mathbb{L}^2(\mathbb{R}^d)$, then $\mathbb{1}_{[-\boldsymbol{k},\boldsymbol{k}]}\widehat{\mathcal{M}}_{\boldsymbol{c}}/\mathcal{M}_{\boldsymbol{c}}[g] \in \mathbb{L}^2(\mathbb{R}^d) \cap \mathbb{L}^1(\mathbb{R}^d)$ and we can define our spectral cut-off density estimator by $\widehat{f}_{\boldsymbol{k}} := \mathcal{M}_{\boldsymbol{c}}^{-1}[\mathbb{1}_{[-\boldsymbol{k},\boldsymbol{k}]}\widehat{\mathcal{M}}_{\boldsymbol{c}}/\mathcal{M}_{\boldsymbol{c}}[g]]$. More explicitly, we have

$$\widehat{f}_{\boldsymbol{k}}(\boldsymbol{x}) = \frac{1}{(2\pi)^d}\int_{[-\boldsymbol{k},\boldsymbol{k}]} \underline{\boldsymbol{x}}^{-\boldsymbol{c}-i\boldsymbol{t}} \frac{\widehat{\mathcal{M}}_{\boldsymbol{c}}(\boldsymbol{t})}{\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})} d\lambda^d(\boldsymbol{t}), \qquad \text{for } \boldsymbol{x} \in \mathbb{R}^d_+. \tag{3.29}$$

Up to now, we had two minor assumptions on the error density g, which we want to collect in the following assumption:

$$\forall \boldsymbol{t} \in \mathbb{R}^d : \mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t}) \neq 0 \text{ and } \forall \boldsymbol{k} \in \mathbb{R}^d_+ : \int_{[-\boldsymbol{k},\boldsymbol{k}]} |\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})|^{-2} d\lambda^d(\boldsymbol{t}) < \infty. \tag{[G0]}$$

The following proposition shows that the proposed estimator is consistent for a suitable choice of the cut-off parameter $\boldsymbol{k} \in \mathbb{R}^d_+$. Its proof is postponed to the proof section, Section 3.5.7.

Proposition 3.5.1 (Upper bound of the risk)**:**
Let $f \in \mathbb{L}^2(\mathbb{R}^d_+, \underline{\boldsymbol{x}}^{2\boldsymbol{c}-\boldsymbol{1}})$, $\sigma := \mathbb{E}_{f_Y}(\underline{\boldsymbol{Y}}^{2\boldsymbol{c}-\boldsymbol{2}}) < \infty$ and assume that **[G0]** holds for g. Then we have for any $\boldsymbol{k} \in \mathbb{R}^d_+$,

$$\mathbb{E}^n_{f_Y}(\|f - \widehat{f}_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}}^{2\boldsymbol{c}-\boldsymbol{1}}}) = \|f - f_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}}^{2\boldsymbol{c}-\boldsymbol{1}}} + \frac{\sigma\Delta_g(\boldsymbol{k})}{n}, \tag{3.30}$$

where $\Delta_g(\boldsymbol{k}) := \frac{1}{(2\pi)^d}\int_{[-\boldsymbol{k},\boldsymbol{k}]} |\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})|^{-2} d\lambda^d(\boldsymbol{t})$. Now choosing $\boldsymbol{k}_n$ such that $\Delta_g(\boldsymbol{k}_n)n^{-1} \to 0$ and $\boldsymbol{k}_n \to \infty$ implies the consistency of $\widehat{f}_{\boldsymbol{k}_n}$.

Let us comment the last result. For a suitable choice of the spectral cut-off parameter $\boldsymbol{k} \in \mathbb{R}^d_+$ we can show that the estimator is consistent in the sense of the weighted $\mathbb{L}^2$ distance.

Example 3.5.2 (Error distribution: Independent variates)**:**
Let $\boldsymbol{U}_1$ have stochastic independent variates, that is $U_{1,1}, \dots, U_{1,d}$ are stochastically independent and $g(\boldsymbol{x}) = \prod_{j\in[\![d]\!]} g_j(x_j)$ for $\boldsymbol{x} \in \mathbb{R}^d_+$. Then the Mellin transform g factorize as seen in Example 2.4.4. For $\boldsymbol{c} \in \mathbb{R}^d$ such that $\mathbb{E}_g(\underline{\boldsymbol{U}}^{\boldsymbol{c}-1}) < \infty$ and $\boldsymbol{k} \in \mathbb{R}^d_+$ holds

$$\Delta_g(\boldsymbol{k}) = \frac{1}{(2\pi)^d}\int_{[-\boldsymbol{k},\boldsymbol{k}]} \prod_{j\in[\![d]\!]} |\mathcal{M}_{c_j}[g_j](t_j)|^{-2} d\lambda^d(\boldsymbol{t}) = \prod_{j\in[\![d]\!]} \Delta_{g_j}(\boldsymbol{k}_j),$$

in other words, Δ_g factorizes, too.

Up to now, the assumptions on f and g were to ensure the well-definedness of the estimator and the weighted $\mathbb{L}^2$-risk. Here, we can already see that the first term, the bias term, in Proposition 3.5.1 is decreasing if $\boldsymbol{k} \in \mathbb{R}_+$ is increasing in any direction while the second summand, called variance term, is increasing. For a more sophisticated analysis of both terms we will consider stronger assumptions on the densities f and g.

3.5.3 Minimax theory

Starting with the noise density g we will present in analogy to Subsection 3.2.3 a multivariate definition of SMOOTH ERROR DENSITIES. We will then make use of the anisotropic Mellin Sobolev spaces, compare 2.4.14, to control the bias term and develop the minimax-optimality of our estimator by stating both, an upper and a lower, bound of the risk uniformly over the anisotropic Mellin-Sobolev ellipsoids.

As already mentioned, the variance term in (3.30) is monotonically increasing in each component $k_j, j \in [\![d]\!]$, of $\boldsymbol{k}$. More precisely, the growth of Δ_g is determined by the decay of the Mellin transform of g in each direction.

Definition 3.5.3 (Multivariate smooth error density)**:**
Let $g : \mathbb{R}_+^d \to \mathbb{R}_+$ such that $\mathbb{E}_g(\boldsymbol{U}^{\underline{\boldsymbol{c}-\boldsymbol{1}}}) < \infty$ for a $\boldsymbol{c} \in \mathbb{R}^d$. Then we call the error density g is a SMOOTH ERROR DENSITY, if there exist $c_g, C_g \in \mathbb{R}_+$ and $\boldsymbol{\gamma} \in \mathbb{R}_+^d$ such that

$$c_g \prod_{j\in[\![d]\!]} (1+t_j^2)^{-\gamma_j/2} \leq |\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})| \leq C_g \prod_{j\in[\![d]\!]} (1+t_j^2)^{-\gamma_j/2} \text{ for all } t \in \mathbb{R}. \qquad \textbf{([G1])}$$

This assumption on the error density was also considered in the work of Belomestny and Goldenshluger (2020) in the univariate case and Comte and Lacour (2013) for multivariate additive deconvolution. Under **[G1]** we see that $\Delta_g(\boldsymbol{k}) \leq C_g \prod_{j\in[\![d]\!]} k_j^{2\gamma_j+1}$ for every $\boldsymbol{k} \in \mathbb{R}_+^d$.

Remark 3.5.4 (Error distribution: Independent variates)**:**
In the case of independent variates, we see that an error density g is smooth with $\boldsymbol{\gamma} \in \mathbb{R}_+^d$ if and only if for each $j \in [\![d]\!]$ the marginal distribution g_j is smooth with parameter $\gamma_j \in \mathbb{R}_+$.

For the sake of completeness, let us now give an example of a smooth error density, without independent variates.

Example 3.5.5 (Error distribution: Dependent variates)**:**
Let $S := (\boldsymbol{0}, \boldsymbol{2}) \setminus (\boldsymbol{0}, \boldsymbol{1})$ and g the density of the uniform distribution on S which is then given by $g(\boldsymbol{x}) = \mathbb{1}_S(\boldsymbol{x})3^{-1}$. The support of the marginal distributions g_1, g_2 of g are both $(0, 2)$ implying that $g_1(x_1)g_2(x_2) \neq g(\boldsymbol{x})$ for all $\boldsymbol{x} \in (\boldsymbol{0}, \boldsymbol{1})$, in other words g does not factorize. Then, for $\boldsymbol{c} = \boldsymbol{1}$ holds

$$\mathcal{M}_{\boldsymbol{1}}[g](\boldsymbol{t}) = \frac{1}{3}\left(\int_{(\boldsymbol{0},\boldsymbol{2})} \boldsymbol{x}^{\underline{i\boldsymbol{t}}} d\boldsymbol{x} - \int_{(\boldsymbol{0},\boldsymbol{1})} \boldsymbol{x}^{\underline{i\boldsymbol{t}}} d\boldsymbol{x}\right) = \frac{2^{2+i(t_1+t_2)} - 1}{3(1+it_1)(1+it_2)}.$$

Further, for all $\boldsymbol{t} \in \mathbb{R}^2$ holds

$$\frac{1}{3} \leq |\mathcal{M}_{\boldsymbol{1}}[g](\boldsymbol{t})|(1+t_1^2)^{1/2}(1+t_2^2)^{1/2} \leq 1,$$

implying that g fulfills **[G1]** with $\boldsymbol{\gamma} = \boldsymbol{1}$.

After a more sophisticated bound of the variance term we will consider now the bias term which occurs in (3.30). Let us for $\boldsymbol{s}, \boldsymbol{c} \in \mathbb{R}_+^d$ consider the ANISOTROPIC MELLIN-SOBOLEV SPACE defined in Definition 2.4.14, namely

$$\mathbb{W}_{\boldsymbol{c}}^{\boldsymbol{s}}(\mathbb{R}_+^d) := \{h \in \mathbb{L}^2(\mathbb{R}_+, \boldsymbol{x}^{\underline{2c-1}}) : |h|_{\boldsymbol{s},\boldsymbol{c}}^2 := \sum_{j \in [\![d]\!]} \|(1+t_j)^{s_j} \mathcal{M}_{\boldsymbol{c}}[h]\|_{\mathbb{R}^d}^2 < \infty\} \tag{3.31}$$

and the corresponding ellipsoids with $L \in \mathbb{R}_+$ $\mathbb{W}_{\boldsymbol{c}}^{\boldsymbol{s}}(L) := \{h \in \mathbb{W}_{\boldsymbol{c}}^{\boldsymbol{s}}(\mathbb{R}_+^d) : |h|_{\boldsymbol{s},\boldsymbol{c}}^2 \leq L\}$. Since $\bigcup_{j=1}^d \{\boldsymbol{t} \in \mathbb{R}^d : |t_j| > k_j\} \supset [-\boldsymbol{k}, \boldsymbol{k}]^c$ we deduce from the assumption $f \in \mathbb{W}_{\boldsymbol{c}}^{\boldsymbol{s}}(L)$ that

$$\|f - f_{\boldsymbol{k}}\|_{\boldsymbol{x}^{\underline{2c-1}}}^2 \leq \sum_{j \in [\![d]\!]} \|\mathbb{1}_{[-k_j,k_j]^c} \mathcal{M}_{\boldsymbol{c}}[f]\|_{\mathbb{R}^d}^2 \leq L \sum_{j \in [\![d]\!]} k_j^{-2s_j}.$$

Setting $\mathbb{D}_{\boldsymbol{c}}^{\boldsymbol{s}}(L) := \{f \in \mathbb{W}_{\boldsymbol{c}}^{\boldsymbol{s}}(L) : f \text{ density}, \mathbb{E}_f(\boldsymbol{X}^{\underline{2c-2}}) \leq L\}$, the previous discussion leads to the following statement whose proof is thus omitted.

Proposition 3.5.6 (Upper bound for the minimax risk)**:**
Let $f \in \mathbb{L}^2(\mathbb{R}_+^d, \boldsymbol{x}^{\underline{2c-1}})$ and $\mathbb{E}_g(\boldsymbol{U}^{\underline{2c-2}}) < \infty$. Then under the assumption **[G1]** it holds

$$\sup_{f \in \mathbb{D}_{\boldsymbol{c}}^{\boldsymbol{s}}(L)} \mathbb{E}_{f_Y}^n (\|f - \widehat{f}_{\boldsymbol{k}_o}\|_{\boldsymbol{x}^{\underline{2c-1}}})^2 \leq C(L, g, \boldsymbol{s}) n^{-1/(1+2^{-1}\sum_{j \in [\![d]\!]}(2\gamma_j+1)s_j^{-1})}$$

for the choice $\boldsymbol{k}_o = (k_{1,o}, \dots, k_{d,o})$ with $k_{i,o} := n^{1/(2s_i + s_i \sum_{j \in [\![d]\!]}(2\gamma_j+1)s_j^{-1})}$.

Before presenting a lower bound of the risk, we illustrate the result presented in Proposition 3.5.6.

Remark 3.5.7:
We consider the following special cases which are derived from Proposition 3.5.6

(i) For the case of isotropic noise, that is $\gamma_1 = \dots = \gamma_d = \gamma \in \mathbb{R}_+$, the resulting rate is given by

$$\sup_{f \in \mathbb{D}_{\boldsymbol{c}}^{\boldsymbol{s}}(L)} \mathbb{E}_{f_Y}^n (\|f - \widehat{f}_{\boldsymbol{k}_o}\|_{\boldsymbol{x}^{\underline{2c-1}}})^2 \leq C(L, g, \boldsymbol{s}) n^{-2\overline{s}/(2\overline{s}+2d\gamma+d)}$$

for $\overline{s} := (\frac{1}{d}\sum_{j \in [\![d]\!]} s_j^{-1})^{-1}$, which is a natural extension to the expected rate in anisotropic density estimation, compare Comte (2017) for the case $d = 2$.

(ii) For the case of isotropic regularity, that is $s_1 = \dots = s_d = s \in \mathbb{R}_+$ we get

$$\sup_{f \in \mathbb{D}_{\boldsymbol{c}}^{\boldsymbol{s}}(L)} \mathbb{E}_{f_Y}^n (\|f - \widehat{f}_{\boldsymbol{k}_o}\|_{\boldsymbol{x}^{\underline{2c-1}}})^2 \leq C(L, g, \boldsymbol{s}) n^{-2s/(2s+2d\overline{\gamma}+d)},$$

where $\overline{\gamma} := d^{-1} \sum_{j [\![d]\!]} \gamma_j$.

(iii) For the case $\gamma_1 = \dots = \gamma_d = \gamma$ and $s_1 = \dots = s_d = s \in \mathbb{R}_+$ we get the usual isotropic rate

$$\sup_{f \in \mathbb{D}_{\boldsymbol{c}}^{\boldsymbol{s}}(L)} \mathbb{E}_{f_Y}^n (\|f - \widehat{f}_{\boldsymbol{k}_o}\|_{\boldsymbol{x}^{\underline{2c-1}}})^2 \leq C(L, g, \boldsymbol{s}) n^{-2s/(2s+2\gamma d+d)}.$$

Now to show that the rate presented in Lemma 3.5.6 is the minimax rate, compare 3.1.3 we are in need of a lower bound result. From this we deduce that our estimator $\widehat{f}_{k_n}$ is minimax-optimal

over the ellipsoids $\mathbb{D}_{\boldsymbol{c}}^{\boldsymbol{s}}(L)$ for many classes of error densities.
For the following part, we will need to have further assumption on the error density g. In fact, we will distinguish if $c_j \in (0, 1/2]$ or $c_j > 1/2$ for $j \in [\![d]\!]$. Let therefore $\tilde{\boldsymbol{c}} := (\tilde{c}_1, \dots, \tilde{c}_d) \in \mathbb{R}_+^d$, where $\tilde{c}_j = 2^{-1}\mathbb{1}_{(1/2,\infty)}(c_j)$. Let us assume that g has a bounded support, that is for all $\boldsymbol{x} \in \mathbb{R}_+^d \setminus (0,1)^d g(\boldsymbol{x}) = 0$ and that there exist constants $c_g, C_g > 0$ such that

$$c_g \prod_{j\in[\![d]\!]} (1+t_j^2)^{-\gamma_j/2} \leq |\mathcal{M}_{\tilde{\boldsymbol{c}}}[g](\boldsymbol{t})| \leq C_g \prod_{j\in[\![d]\!]} (1+t_j^2)^{-\gamma_j/2} \text{ for all } \boldsymbol{t} \in \mathbb{R}^d. \qquad \textbf{([G1'])}$$

With this additional assumption we can show the following theorem where its proof can be found in Subsection 3.5.7.

Theorem 3.5.8 (Lower bound for the minimax risk):
Let $\boldsymbol{s}, \boldsymbol{\gamma} \in \mathbb{N}^d$, $\boldsymbol{c} > \boldsymbol{0}$ and assume that **[G1]** and **[G1']** hold. Then there exist constants $C_{g,c}, L_{\boldsymbol{s},g,\boldsymbol{c}} > 0$, such that for all $L \geq L_{\boldsymbol{s},g,\boldsymbol{c}}$, $n \in \mathbb{N}$ and for any estimator $\widehat{f}$ of f based on an i.i.d. sample $(\boldsymbol{Y}_j)_{j\in[\![n]\!]}$,

$$\sup_{f\in\mathbb{D}_{\boldsymbol{c}}^{\boldsymbol{s}}(L)} \mathbb{E}_{f_Y}^n(\|\widehat{f}-f\|^2_{\boldsymbol{x}^{\underline{2\boldsymbol{c}-\boldsymbol{1}}}}) \geq C_{g,\boldsymbol{c}} n^{-1/(1+2^{-1}\sum_{j\in[\![d]\!]}(2\gamma_j+1)s_j^{-1})}.$$

Let us shortly comment on the assumption **[G1']**. For the case that $c_j > 1/2$ for $j \in [\![d]\!]$ then $\tilde{c}_j = 1/2$, and hence we need to assume that $\mathbb{E}_g(U_j^{-1/2}) < \infty$, which is a mild condition. If $c_j \leq 1/2$ we have that $\tilde{c}_j = 0$. In this case, we have automatically that $\mathbb{E}_g(U_j^{-1}) < \infty$, if we assume that $\mathbb{E}_g(U_j^{2c_j-2}) < \infty$, compare Proposition 3.5.1.

3.5.4 Data-driven method

Although we have shown that in certain situations the estimator $\widehat{f}_{\boldsymbol{k}_n}$ in Proposition 3.5.6 is minimax-optimal, the choice of $\boldsymbol{k}_n$ is still dependent on the regularity parameter $\boldsymbol{s} \in \mathbb{R}_+^d$ of the unknown density f, which is again, unknown. Therefore, we will propose a fully data-driven choice of $\boldsymbol{k} \in \mathbb{R}_+^d$ based on the sample $(\boldsymbol{Y}_j)_{j\in[\![n]\!]}$. To do so, we will use a model selection approach. For the special case of $d = 1$, we proposed a data-driven choice for the parameter $k \in \mathbb{R}_+$ based on a penalized contrast approach in Subsection 3.2.4. For the multivariate case, a model selection approach has been mainly used, if one considers an isotropic choice of the cut-off parameter; that is, instead of considering the estimator defined in (3.29) one would use for $k \in \mathbb{R}_+$ and $[-k,k]^d := \{\boldsymbol{x} \in \mathbb{R}^d : \forall j \in [\![d]\!] : |x_j| \leq k\}$ the estimator

$$\widetilde{f}_k(\boldsymbol{x}) = \frac{1}{(2\pi)^d} \int_{[-k,k]^d} \boldsymbol{x}^{\underline{-\boldsymbol{c}-i\boldsymbol{t}}} \frac{\widehat{\mathcal{M}}_{\boldsymbol{c}}(\boldsymbol{t})}{\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})} d\boldsymbol{t}, \quad \boldsymbol{x} \in \mathbb{R}_+^d.$$

For the family $(\widetilde{f}_k)_{k\in\mathbb{R}_+}$ a data-driven choice of the parameter $k \in \mathbb{R}_+$ based on a model selection approach is possible. For the anisotropic estimator defined in (3.29) we propose a data-driven choice based on a model selection, which can be used even for anisotropic choices of the cut-off parameter $\boldsymbol{k} \in \mathbb{R}_+^d$. To the knowledge of the author the usage of a model selection approach instead of a Lepski approach, Dussap (2021); Comte and Lacour (2013), has not been considered so far.
Let us reduce the set of possible parameters to

$$\mathcal{K}_n := \{\boldsymbol{k} \in \mathbb{N}^d : \Delta_g(\boldsymbol{k}) \leq n\},$$

and define for $\boldsymbol{k} \in \mathbb{R}_+^d$ and $\chi > 0$ the penalty term

$$\mathrm{pen}(\boldsymbol{k}) := \chi\sigma\Delta_g(\boldsymbol{k})n^{-1}.$$

It can be seen that the bias $\|f - f_{\boldsymbol{k}}\|^2_{\boldsymbol{x}^{\underline{2c-1}}} = \|f\|^2_{\boldsymbol{x}^{\underline{2c-1}}} - \|f_{\boldsymbol{k}}\|^2_{\boldsymbol{x}^{\underline{2c-1}}}$ behaves like $-\|f_{\boldsymbol{k}}\|^2_{\boldsymbol{x}^{\underline{2c-1}}}$. Exchanging $-\|f_{\boldsymbol{k}}\|^2_{\boldsymbol{x}^{\underline{2c-1}}}$ and $\mathrm{pen}(\boldsymbol{k})$ with their empirical counterparts, we define $\widehat{\boldsymbol{k}}$ by the following expression:

$$\widehat{\boldsymbol{k}} := \arg\min_{\boldsymbol{k}\in\mathcal{K}_n}(-\|\widehat{f}_{\boldsymbol{k}}\|^2_{\boldsymbol{x}^{\underline{2c-1}}} + \widehat{\mathrm{pen}}(\boldsymbol{k})), \quad \widehat{\mathrm{pen}}(\boldsymbol{k}) := \chi\widehat{\sigma}\Delta_g(\boldsymbol{k})n^{-1}, \tag{3.32}$$

where $\widehat{\sigma} := n^{-1}\sum_{j\in[\![n]\!]}\boldsymbol{Y}_j^{\underline{2c-2}}$. Now, let us show that this data-driven procedure mimics the optimal choice up to a negligible term.

Theorem 3.5.9 (Data-driven anisotropic choice of $\boldsymbol{k} \in \mathbb{R}_+^d$)**:**
Assume that $\mathbb{E}_{f_{\boldsymbol{Y}}}(\boldsymbol{Y}^{\underline{7(c-1)}}) < \infty$, $\|\boldsymbol{x}^{\underline{2c-1}}f_{\boldsymbol{Y}}\|_\infty < \infty$ and **[G1]** is fulfilled. Then, for $\chi \geq 144$ we have

$$\mathbb{E}^n_{f_{\boldsymbol{Y}}}(\|f - \widehat{f}_{\widehat{\boldsymbol{k}}}\|^2_{\boldsymbol{x}^{\underline{2c-1}}}) \leq 3\inf_{\boldsymbol{k}\in\mathcal{K}_n}(\|f - f_{\boldsymbol{k}}\|^2_{\boldsymbol{x}^{\underline{2c-1}}} + \mathrm{pen}(\boldsymbol{k})) + \frac{C_2}{n},$$

$C_2 > 0$ is a constant depending on χ, $\|f_{\boldsymbol{Y}}\boldsymbol{x}^{\underline{2c-1}}\|_\infty$, σ,$\mathbb{E}_{f_{\boldsymbol{Y}}}(\boldsymbol{Y}^{\underline{7(c-1)}})$ and g.

For every $\boldsymbol{s} \in \mathbb{R}_+^d$ we can see that $\boldsymbol{k}_n$, defined in Proposition 3.5.6 lies in $\mathcal{K}_n$. Due to this, and the consideration in the minimax theory section, we can deduce the following Corollary directly, whose proof is thus omitted.

Corollary 3.5.10:
Under the assumption of Theorem 3.5.9 and the additional assumption that $f \in \mathbb{D}^{\boldsymbol{s}}_{\boldsymbol{c}}(L)$, we get

$$\mathbb{E}^n_{f_{\boldsymbol{Y}}}(\|f - \widehat{f}_{\widehat{\boldsymbol{k}}}\|^2_{\boldsymbol{x}^{2c-1}}) \leq Cn^{-1/(1+2^{-1}\sum_{j\in[\![d]\!]}(2\gamma_j+1)s_j^{-1})},$$

where C is a positive constant depending on χ, $\|f_{\boldsymbol{Y}}\boldsymbol{x}^{\underline{2c-1}}\|_\infty$, σ, $\mathbb{E}_{f_{\boldsymbol{Y}}}(\boldsymbol{Y}^{\underline{7(c-1)}})$, g and L.

3.5.5 Numerical results

Next, we demonstrate the performance of the estimator $\widehat{f}_{\widehat{k}}$ defined in (3.29) and (3.32). We will restrict ourselves to the case $d = 2$. In the bivariate case, we will study the performance of the fully data-driven method presented in (3.32), while we omit the consideration of different values of $\boldsymbol{c} \in \mathbb{R}^2$.

In the upcoming simulation study we will consider the densities

(i) Gamma Distribution: $f_1(x) = \frac{x^3}{96}\exp(-0.5x)\mathbb{1}_{\mathbb{R}_+}(x)$,

(ii) Weibull Distribution: $f_2(x) = 2x\exp(-x^2)\mathbb{1}_{\mathbb{R}^+}(x)$,

(iii) Beta Distribution: $f_3(x) = \frac{1}{560}(0.5x)^3(1-0.5x)^4,1_{[0,1]}(0.5x)$ and

(iv) Log-Normal Distribution: $f_4(x) = \frac{1}{\sqrt{2\pi x}}\exp(-(\log(x)^2/2))\mathbb{1}_{(0,\infty)}(x)$.

For the error densities, we consider the univariate densities

(i) Pareto Distribution: $g_1(x) = x^{-2}\mathbb{1}_{(1,\infty)}(x)$, see Tables 3.1,

(ii) Log-Gamma Distribution: $g_2(x) = \frac{1}{\Gamma(1/2)}x^{-2}\log(x)^{-1/2}\mathbb{1}_{(1,\infty)}(x)$.

In the sense of **[G1]** we see that g_1 has the parameter $\gamma = 1$, while g_2 has $\gamma = 1/2$. Let us now consider the data-driven choice defined (3.29) and (3.32) for the case of $d = 2$. To illustrate the performance of our estimator, we consider the following four cases

(i) Error densities: $f(x_1, x_2) = f_1(x_1)f_1(x_1)$ with direct observations compared to observations with $g(x_1, x_2) = g_2(x_1)g_2(x_2)$ with $\boldsymbol{c} = (1/2, 1/2)^T$,

(ii) Anisotropic density: $f(x_1, x_2) = f_3(x_1)f_4(x_2)$ with direct observations and $c = (1/2, 1/2)^T$,

(iii) Dependency: $\boldsymbol{X} \sim \mathrm{LN}_{\mu,\boldsymbol{\Sigma}}$ with $\boldsymbol{\mu} = (\log(4), \log(4))^T$, $c = (1/2, 1/2)^T$ and direct observations. For $\boldsymbol{\Sigma}$ we compare

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0.81 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 0.81 \end{pmatrix},$$

and

(iv) Anisotropic error: $f(\boldsymbol{x}) = f_2(x_1)f_2(x_2)$ with $\boldsymbol{U} = (U_1, U_2)^T$, where $U_1 \sim \mathrm{L}\Gamma_{(1,1)}$ and $U_2 \sim \mathrm{L}\Gamma_{(1/2,1)}$, dependent, and $\boldsymbol{c} = (1/2, 1/2)^T$, compare Example 2.1.3.

For the first case, we visualize the impact of observations with measurement error compared to direct observations. The second case resembles the case when the decay of the Mellin transform of the density f has significantly different behavior in different directions. The third case will illustrate the behavior of the estimator when the two coordinates of X are dependent, and in the fourth case the decay of the Mellin transform of the density is similar but the decays of the error density are not the same. By minimizing an integrated weighted squared error over a family of histogram densities with randomly drawn partitions and weights, we select $\chi = 1$ (respectively $\chi = 0.3$) for the cases of direct observation (respectively contaminated data), where χ is the variance constant, see (3.32).

Case 1: We begin by examining the influence of measurement errors by comparing the estimator $\widehat{f}_{\widehat{\boldsymbol{k}}}$ based on the copies of $\boldsymbol{X}$ compared to copies of $\boldsymbol{Y}$.

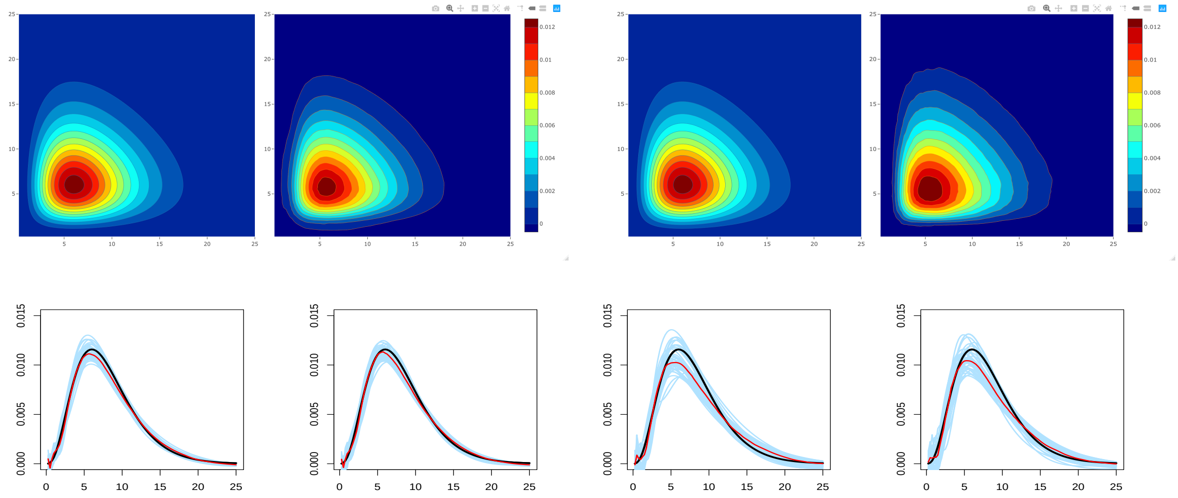

Figure 3.13: The estimator $\widehat{f}_{\widehat{\boldsymbol{k}}}$ are depicted for 50 Monte-Carlo simulations with sample size $n = 1000$ with direct observations (left) and with multiplicative measurement errors (right). The top plots are the true density (left) and the point wise median of the estimators (right). The bottom plots are the sections for $x = 7.5$ (right) and $y = 7.5$ (left), where the true density f is given by the black curve while the red curve is the point wise empirical median of the 50 estimates.

Case 2: In the second case, we additionally compare the anisotropic estimator $\widehat{f}_{\widehat{\boldsymbol{k}}}$ with the isotropic choice, that is we define $\widehat{f}_{\widetilde{k}} := \widehat{f}_{\widetilde{k},\widetilde{k}}$ with

$$\widetilde{k} := \arg \min_{k\in\mathbb{N}, \Delta_g((k,k))\leq n} -\|\widehat{f}_{(k,k)}\|^2_{x^{2c-1}} + \kappa\widehat{\sigma}\frac{\Delta_g((k,k))}{n}, \tag{3.33}$$

a penalized contrast approach which is a direct generalization of the estimator given in Section 3.2. Here we choose $\kappa = 5$ by a preliminary simulation study.

As we see in Fig. 3.14, the anisotropic estimator seems to invest more in the approximation of the Beta distribution than the Log normal distribution. This leads to worse performance in the Log normal direction but to an overall satisfying result. In comparison to that, the isotropic estimator chooses in both direction the same cut-off parameter leading to a better approximation of the log normal distribution but also to a worse approximation of the beta distribution. Overall it seems that the anisotropic estimator behaves better.

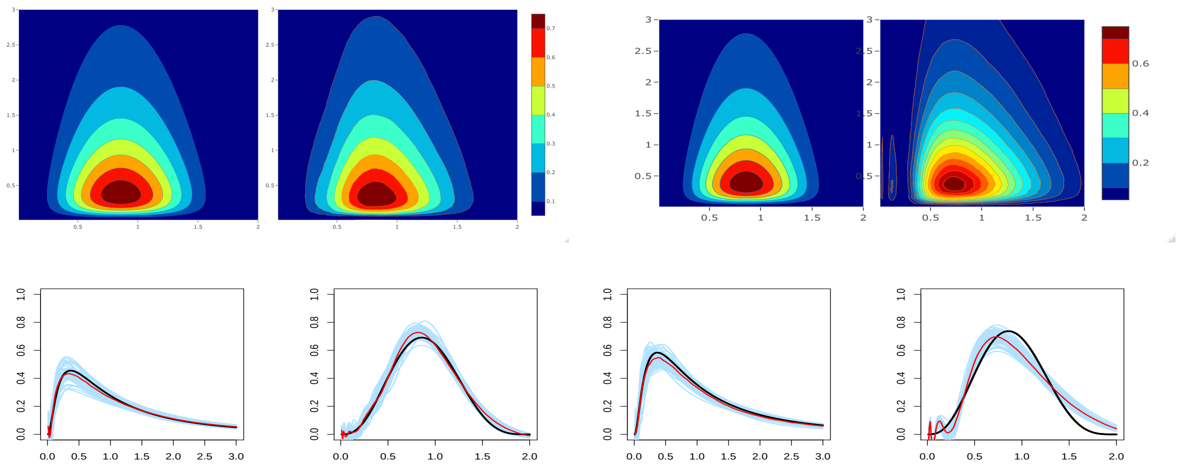

Figure 3.14: The estimator $\widehat{f}_{\widehat{\boldsymbol{k}}}$ (left) and $\widehat{f}_{\widetilde{\boldsymbol{k}}}$ (right) are depicted for 50 Monte-Carlo simulations with sample size $n = 1000$ with direct observations. The top plots are the true density (left) and the point wise median of the estimators (right). The bottom plots are the sections for $x = 5$ (right) and $y = 0.59$ (left) where the true density f is given by the black curve while the red curve is the point wise empirical median of the 50 estimates.

Case 3: Now we consider the influence of the dependency between the coordinates of $\boldsymbol{X}$. While $\boldsymbol{\Sigma}_1$ resembles the case of independent coordinates, $\boldsymbol{\Sigma}_2$ is not a diagonal matrix and thus the coordinates are dependent.

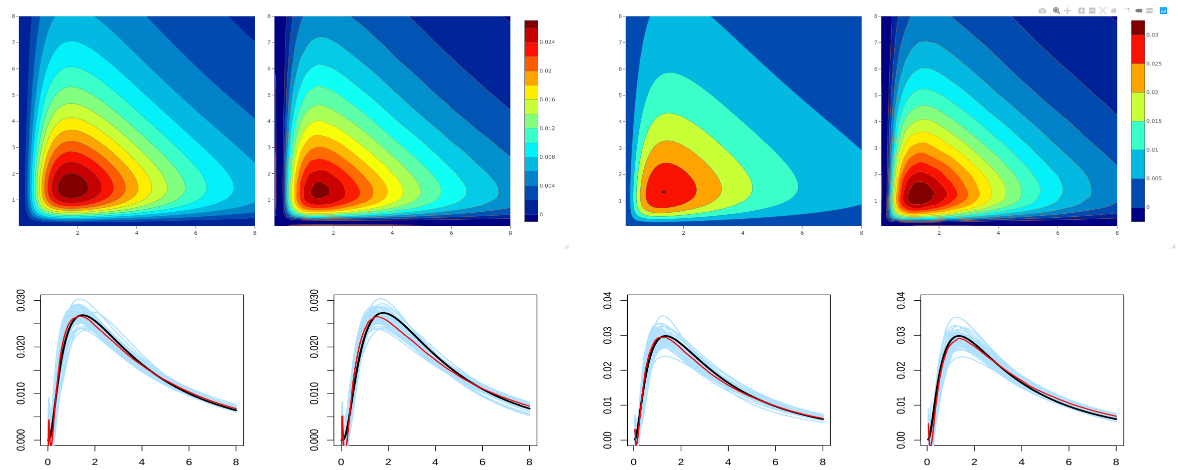

Figure 3.15: The estimator $\widehat{f}_{\widehat{\boldsymbol{k}}}$ are depicted for 50 Monte-Carlo simulations with sample size $n = 1000$ with direct observations and $\boldsymbol{\Sigma}_1$ (left) and $\boldsymbol{\Sigma}_2$ (right). The top plots are the true density (left) and the point wise median of the estimators (right). The bottom plots are the sections for $x = 1.5$ (right) and $y = 1.5$ (left), where the true density f is given by the black curve while the red curve is the point wise empirical median of the 50 estimates.

In Fig. 3.15 we can see that although the estimator does reconstruct the general shape of the density

f, the included dependency slightly impedes the estimation.

Case 4: We finish our simulation study by considering the case, where the decay of the Mellin transform of the density behaves similar in both direction, while the decay of the Mellin transform of the error densities differs. For $\boldsymbol{U} = (U_1, U_2)^T$ we set $U_1 := \xi_1 \xi_2$ and $U_2 := \xi_2$ where $\xi_1, \xi_2 \overset{i.i.d.}{\sim} \mathrm{L}\Gamma_{(1/2,1)}$, compare Example 2.1.3. Then we have $U_1 \sim \mathrm{L}\Gamma_{(1,1)}$, $U_2 \sim \mathrm{L}\Gamma_{(1/2,1)}$ and

$$\mathcal{M}_{\mathbf{1/2}}[g](\boldsymbol{t}) = (3/2 - it_1)^{-1/2}(1 - i(t_1 + t_2))^{-1/2}, \quad \boldsymbol{t} \in \mathbb{R}^2,$$

leading to that g satisfies in this situation **[G1]** with $\boldsymbol{\gamma} = (1, 1/2)$. For this case, we found by a preliminary simulation study the choice $\chi_1 = \chi_2 = 0.25$. For the distribution of $\boldsymbol{X} = (X_1, X_2)^t$ we set for the sake of simplicity $X_1, X_2 \overset{i.i.d.}{\sim} f_2$. We compare the performance of the data-driven anisotropic estimator $\widehat{f}_{\widetilde{\boldsymbol{k}}}$ with the performance of the data-driven isotropic estimator $\widehat{f}_{\widetilde{k}}$ introduced in (3.33) with the choice $\kappa = 1.02$. For both estimator we consider $\boldsymbol{c} = \mathbf{1/2} = (1/2, 1/2)^T$.

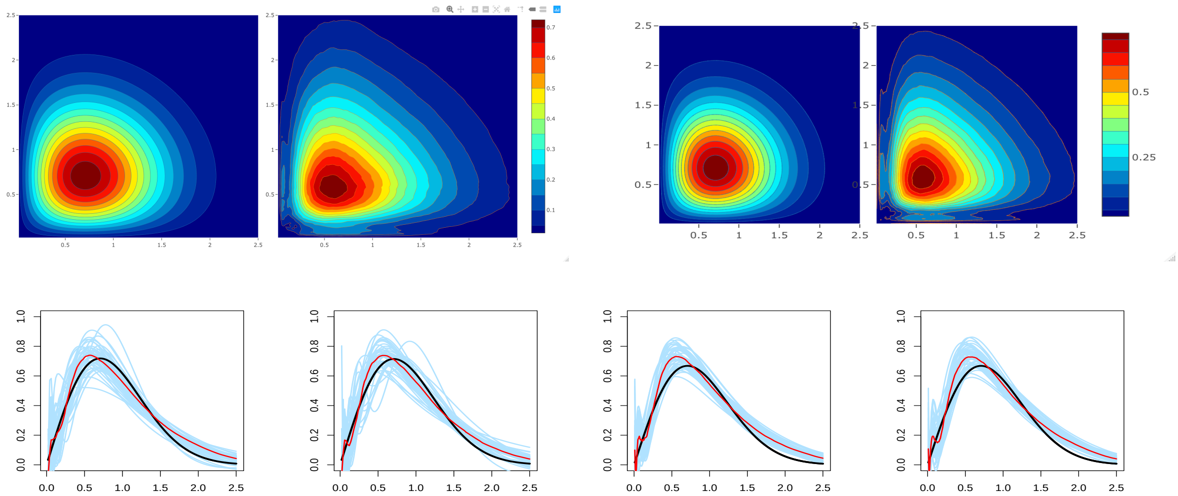

Figure 3.16: The estimator $\widehat{f}_{\widetilde{\boldsymbol{k}}}$ (left) and $\widehat{f}_{\widetilde{k}}$ (right) are depicted for 50 Monte-Carlo simulations with sample size $n = 1000$ with observations under multiplicative measurement errors. The top plots are the true density (left) and the point wise median of the estimators (right). The bottom plots are the sections for $x = 5$ (right) and $y = 0.59$ (left), where the true density f is given by the black curve while the red curve is the point wise empirical median of the 50 estimates.

As we see in Fig. 3.16 the anisotropic estimator $\widehat{f}_{\widetilde{k}}$ behaves better in the second coordinates as the isotropic estimator $\widehat{f}_{\widetilde{k}}$, which is consistent with the theory since the decay of the Mellin transform of the error density in this direction is slower.

3.5.6 Conclusion

In this section, we generalized the results from Section 3.2 for the multivariate case. More precisely, we considered an anisotropic version of the estimator in Section 3.2 and showed its minimax optimality over the anisotropic Mellin Sobolev spaces and proposed a fully data-driven choice of the upcoming spectral cut-off parameter $\boldsymbol{k} \in \mathbb{R}^d_+$. Furthermore, we demonstrated the reasonable behavior of our estimator via a Monte-Carlo simulation study and compared it with its isotropic counterpart.

3.5.7 Proofs

Proof of Proposition 3.5.1. For $\boldsymbol{k} \in \mathbb{R}_+^d$ we see that $f - f_{\boldsymbol{k}} \in \mathbb{L}^2(\mathbb{R}_+^d, \underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}})$ with $\mathcal{M}_{\boldsymbol{c}}[f - f_{\boldsymbol{k}}] = \mathcal{M}_{\boldsymbol{c}}[f]\mathbb{1}_{\mathbb{R}_+^d \setminus [-\boldsymbol{k},\boldsymbol{k}]}$. We deduce by application of the Plancherel identity that $\langle f - f_{\boldsymbol{k}}, f_{\boldsymbol{k}} - \widehat{f}_{\boldsymbol{k}} \rangle_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} = \langle \mathcal{M}_{\boldsymbol{c}}[f]\mathbb{1}_{\mathbb{R}_+^d \setminus [-\boldsymbol{k},\boldsymbol{k}]}, (\mathcal{M}_{\boldsymbol{c}}[f] - \widehat{\mathcal{M}}_{\boldsymbol{c}})\mathbb{1}_{[-\boldsymbol{k},\boldsymbol{k}]} \rangle_{\mathbb{R}^d} = 0$, which implies that

$$\mathbb{E}_{f_Y}^n(\|f - \widehat{f}_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}}) = \|f - f_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} + \mathbb{E}_{f_Y}^n(\|\widehat{f}_{\boldsymbol{k}} - f_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}}).$$

Now by application of the Fubini-Tonelli theorem we interchange the integration order to get

$$\begin{aligned}\mathbb{E}_{f_Y}^n(\|\widehat{f}_{\boldsymbol{k}} - f_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}}) &= \frac{1}{(2\pi)^d} \int_{[-\boldsymbol{k},\boldsymbol{k}]} \mathbb{E}_{f_Y}^n(|\mathcal{M}_c[f_Y](\boldsymbol{t}) - \widehat{\mathcal{M}}_{\boldsymbol{c}}(\boldsymbol{t})|^2)|\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})|^{-2} d\lambda^d(\boldsymbol{t}) \\ &\leq \frac{1}{(2\pi)^d n}\sigma\Delta_g(\boldsymbol{k}).\end{aligned}$$

□

Proof of Theorem 3.5.8. First we outline the main steps of the proof. Let us denote by $\mathcal{I} := \{j \in [\![d]\!] : c_j > 1/2\}$ the subset of indices. We will construct a family of functions in $\mathbb{D}_{\boldsymbol{c}}^{\boldsymbol{s}}(L)$ by a perturbation of the density $f_o : \mathbb{R}_+ \to \mathbb{R}_+$ with small bumps, such that their $\mathbb{L}^2(\mathbb{R}_+^d, \boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}})$-distance and the Kullback-Leibler divergence of their induced distributions can be bounded from below and above, respectively. The claim then follows then by applying Theorem 2.5 in Tsybakov (2009). We use the following construction, which we present first.

Denote by $C_0^\infty(\mathbb{R})$ the set of all smooth functions with compact support in $\mathbb{R}$ and let $\psi \in C_0^\infty(\mathbb{R})$ be a function with support in $(0,1)$ and $\int_{\mathbb{R}} \psi(x) d\lambda(x) = 0$. For each $j \in [\![d]\!]$, $K_j \in \mathbb{N}$ (to be selected below) and $k_j \in [\![0, K_j - 1]\!]$ we define the bump-functions $\psi_{k_j,K_j}(x_j) := \psi(x_j K_j - K_j - k_j)$, $x_j \in \mathbb{R}$ and define for $p \in \mathbb{N}_0 := \{z \in \mathbb{Z} : z \geq 0\}$ the finite constant $C_{p,\infty} := \max(\|\psi^{(l)}\|_\infty, l \in [\![0,p]\!])$. Let us further use the operator $\mathrm{S} : C_0^\infty(\mathbb{R}) \to C_0^\infty(\mathbb{R})$ with $\mathrm{S}[f](x) = -xf^{(1)}(x)$ for all $x \in \mathbb{R}$ and define $\mathrm{S}^1 := \mathrm{S}$ and $\mathrm{S}^n := \mathrm{S} \circ \mathrm{S}^{n-1}$ for $n \in \mathbb{N}, n \geq 2$. Now, for $p \in \mathbb{N}$, we define the function $\psi_{k_j,K_j,p}(x_j) := \mathrm{S}^p[\psi_{k_j,K_j}](x_j) = \sum_{i=1}^p c_{i,p} x_j^i K_j^i \psi^{(i)}(x_j K_j - K_j - k_j)$ for $x_j \in \mathbb{R}_+$ and $c_{i,p} \geq 1$ and let $c_p := \sum_{i=1}^p c_{i,p}$.
For a bump-amplitude $\delta > 0, \boldsymbol{\gamma} \in \mathbb{N}^d$ and $\boldsymbol{K} := (K_1, \dots, K_d)^T \in \mathbb{N}^d$ define

$$\mathcal{K} := \bigtimes_{j \in [\![d]\!]} [\![0, K_j - 1]\!] := [\![0, K_1 - 1]\!] \times \cdots \times [\![0, K_d - 1]\!] = \{\boldsymbol{k} \in \mathbb{N}_0^d : \forall j \in [\![d]\!] : k_j < K_j\}$$

and for $\boldsymbol{\theta} = (\theta_{\boldsymbol{k}+\mathbf{1}})_{\boldsymbol{k} \in \mathcal{K}} \in \{0,1\}^{\times_{j \in [\![d]\!]} [\![K_j]\!]} =: \boldsymbol{\Theta}$ set

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = f_o(\boldsymbol{x}) + \delta \mathrm{F}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}}^{-1/2} \sum_{\boldsymbol{k}=(k_1,\dots,k_d)^T \in \mathcal{K}} \theta_{\boldsymbol{k}+\mathbf{1}} \prod_{j \in [\![d]\!]} \psi_{k_j,K_j,\gamma_j}(x_j), \tag{3.34}$$

where $\mathrm{F}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}} := \boldsymbol{K}^{\underline{2\gamma}} \sum_{j \in [\![d]\!]} K_j^{2s_j}$ and $f_o(\boldsymbol{x}) := \prod_{j \in [\![d]\!]} f_{o,j}(x_j)$ with

$$f_{o,j}(x) := \begin{cases} \exp(-x)\mathbb{1}_{\mathbb{R}_+}(x), & j \in \mathcal{I}; \\ x\exp(-x)\mathbb{1}_{\mathbb{R}_+}(x), & \text{else.} \end{cases}$$

Until now, we did not give a sufficient condition to ensure that our constructed functions $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ are in fact densities. This condition is given by the following lemma.

Lemma 3.5.11:
Let $0 < \delta < \delta_o(\psi, \boldsymbol{\gamma}) := \exp(-2d)/(\prod_{j\in[\![d]\!]} 2^{\gamma_j} C_{\gamma_j,\infty} c_{\gamma_j})$. Then for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $f_{\boldsymbol{\theta}}$ is a density.

Further, we show that these densities all lie inside the ellipsoids $\mathbb{D}^{\boldsymbol{s}}_{\boldsymbol{c}}(L)$ for L big enough. This is captured in the following lemma.

Lemma 3.5.12:
Let $\boldsymbol{s} \in \mathbb{N}^d$. Then, there is $L_{\boldsymbol{s},\boldsymbol{\gamma},\boldsymbol{c},\delta} > 0$, such that f_o and any $f_{\boldsymbol{\theta}}$ as in (3.34) with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, belong to $\mathbb{D}^{\boldsymbol{s}}_{\boldsymbol{c}}(L_{\boldsymbol{s},\boldsymbol{\gamma},\boldsymbol{c},\delta})$.

For sake of simplicity we denote for a function $\varphi \in \mathbb{L}^2(\mathbb{R}^d_+, \boldsymbol{x}^{\underline{2\tilde{c}-1}})$ the multiplicative convolution with g by $\widetilde{\varphi} := (\varphi * g)$. Further we see that for $\boldsymbol{y} \in (0,2)^d$ it holds

$$\begin{aligned}\widetilde{f}_o(\boldsymbol{y}) &= \boldsymbol{y}^{\underline{1-2\tilde{c}}} \int_{\mathbb{R}^d_+} g(\boldsymbol{x})\boldsymbol{x}^{\underline{2\tilde{c}}} \prod_{j\in[\![d]\!]} \exp(-y_j/x_j) d\lambda^d(\boldsymbol{x}) \\ &\geq \boldsymbol{y}^{\underline{1-2\tilde{c}}} \int_{\mathbb{R}^d_+} g(\boldsymbol{x})\boldsymbol{x}^{\underline{2\tilde{c}}} \prod_{j\in[\![d]\!]} \exp(-2/x_j) d\lambda^d(\boldsymbol{x}) =: c_g \boldsymbol{y}^{\underline{1-2\tilde{c}}}, \end{aligned} \tag{3.35}$$

where $c_g > 0$ since otherwise $g = 0$ almost everywhere. Exploiting Varshamov-Gilbert's lemma (see Tsybakov (2009)) in Lemma 3.5.13 we show further that there is $M \in \mathbb{N}$ with $M \geq 2^{\prod_{j\in[\![d]\!]} K_j/8}$ and a subset $\{\boldsymbol{\theta}^{(0)}, \dots, \boldsymbol{\theta}^{(M)}\}$ of $\boldsymbol{\Theta}$ with $\boldsymbol{\theta}^{(0)} = (0, \dots, 0)$ such that for all $j, l \in [\![0, M]\!]$, $j \neq l$ the $\mathbb{L}^2(\mathbb{R}^d_+, \boldsymbol{x}^{\boldsymbol{2c-1}})$-distance and the Kullback-Leibler divergence are bounded for $\boldsymbol{K} \geq \boldsymbol{K}_o(\boldsymbol{\gamma}, \boldsymbol{c}, \psi)$.

Lemma 3.5.13:
Let $\boldsymbol{K} \geq \boldsymbol{K}_o(\psi, \boldsymbol{\gamma}, \boldsymbol{c})$ (understood component wise). Then there exists a subset $\{\boldsymbol{\theta}^{(0)}, \dots, \boldsymbol{\theta}^{(M)}\}$ of $\boldsymbol{\Theta}$ with $\boldsymbol{\theta}^{(0)} = (0, \dots, 0)$, such that $M \geq 2^{8^{-1}\prod_{j\in[\![d]\!]} K_j}$ and for all $j, l \in [\![0, M]\!], j \neq l$ it holds

$$\begin{aligned} &(i) \quad \|f_{\boldsymbol{\theta}^{(j)}} - f_{\boldsymbol{\theta}^{(l)}}\|^2_{\boldsymbol{x}^{\boldsymbol{2c-1}}} \geq \frac{C_{\boldsymbol{\gamma},\boldsymbol{c}}\delta^2}{\sum_{j\in[\![d]\!]} K_j^{2s_j}} \text{ and} \\ &(ii) \quad \mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}^{(j)}}, \widetilde{f}_{\boldsymbol{\theta}^{(0)}}) \leq \frac{C_{g,\boldsymbol{\gamma}}\delta^2 \log(M) \boldsymbol{K}^{\underline{-2\gamma-1}}}{\sum_{j\in[\![d]\!]} K_j^{2s_j}}, \end{aligned}$$

where KL is the Kullback-Leibler-divergence.

Selecting $K_j = \lceil n^{1/(2s_j + s_j \sum_{i\in[\![d]\!]}(2\gamma_i+1)s_i^{-1}}\rceil$, it follows that

$$\Big(\sum_{j\in[\![d]\!]} K_j^{2s_j}\Big)^{-1} \geq cn^{-1/(1+0.5\sum_{i\in[\![d]\!]}(2\gamma_i+1)s_i^{-1})} \text{ and} \frac{\boldsymbol{K}^{\underline{-2\gamma-1}}}{\sum_{j\in[\![d]\!]} K_j^{2s_j}} \leq n^{-1}$$

for $n \geq n_{\boldsymbol{s},\boldsymbol{\gamma}}$, and thus

$$\frac{1}{M}\sum_{j=1}^M \mathrm{KL}((\widetilde{f}_{\boldsymbol{\theta}^{(j)}})^{\otimes n}, (\widetilde{f}_{\boldsymbol{\theta}^{(0)}})^{\otimes n}) = \frac{n}{M}\sum_{j=1}^M \mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}^{(j)}}, \widetilde{f}_{\boldsymbol{\theta}^{(0)}}) \leq c_{\delta,g,\boldsymbol{\gamma}} \log(M),$$

where $c_{\delta,g,\boldsymbol{\gamma}} < 1/8$ for all $\delta \leq \delta_1(g, \boldsymbol{\gamma}, \boldsymbol{s})$ and $M \geq 2$ for $n \geq n_{\boldsymbol{s},\boldsymbol{\gamma}}$. Thereby, we can use Theorem 2.5 of Tsybakov (2009), which in turn for any estimator $\widehat{f}$ of f implies

$$\sup_{f \in \mathbb{D}_{\boldsymbol{c}}^{\boldsymbol{s}}(L)} \mathbb{P}_{f_{\boldsymbol{Y}}}^n \left(\|\widehat{f} - f\|_{\boldsymbol{x}^{2\boldsymbol{c}-\boldsymbol{1}}}^2 \geq \frac{c_{\delta,\boldsymbol{\gamma},\boldsymbol{c}}}{2} n^{-1/(1+0.5\sum_{i\in[\![d]\!]}(2\gamma_i+1)s_i^{-1})} \right) \geq 0.07.$$

Note that the constant $c_{\delta,\boldsymbol{\gamma},\boldsymbol{c}}$ does only depend on ψ, γ and δ, hence it is independent of the parameters s, L and n. The claim of Theorem 3.5.8 follows by using Markov's inequality, which completes the proof. □

Proof of Lemma 3.5.11. For any $h \in C_0^\infty(\mathbb{R})$ we can state that $\int_{\mathbb{R}} \mathrm{S}[h](x) d\lambda(x) = [-xh(x)]_{-\infty}^{\infty} + \int_{\mathbb{R}} h(x) d\lambda(x) = \int_{\mathbb{R}} h(x) d\lambda(x)$ and therefore $\int_{\mathbb{R}} \mathrm{S}^p[h](x) d\lambda(x) = \int_{\mathbb{R}} h(x) d\lambda(x)$ for $p \in \mathbb{N}$. Thus for every $j \in [\![d]\!]$ we get $\int_{\mathbb{R}} \psi_{k_j,K_j,\gamma_j}(x_j) d\lambda(x_j) = \int_{\mathbb{R}} \psi_{k_j,K_j}(x_j) d\lambda(x_j) = 0$, which implies that for any $\delta > 0$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ we have $\int_{\mathbb{R}_+^d} f_{\boldsymbol{\theta}}(\boldsymbol{x}) d\lambda^d(\boldsymbol{x}) = 1$.
Now due to the construction (3.34) of the functions ψ_{k_j,K_j} we easily see that the function ψ_{k_j,K_j} has support on $[1 + k_j/K_j, 1 + (k_j+1)/K_j]$, which leads to ψ_{k_j,K_j} and ψ_{l_j,K_j} having disjoint supports if $k_j \neq l_j$. Here, we want to emphasize that $\mathrm{supp}(\mathrm{S}[h]) \subseteq \mathrm{supp}(h)$ for all $h \in C_0^\infty(\mathbb{R})$. This implies that ψ_{k_j,K_j,γ_j} and ψ_{l_j,K_j,γ_j} have disjoint supports if $k_j \neq l_j$, too. For $\boldsymbol{x} \in \mathbb{R}_+^d \setminus \times_{j\in[\![d]\!]}[1,2]$ we have $f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \exp(-\sum_{j\in[\![d]\!]} x_j) \prod_{j\in\mathcal{I}^c} x_j \geq 0$. Now let us consider the case $\boldsymbol{x} \in \times_{j\in[\![d]\!]}[1,2]$. In fact there are $k_{1,o} \in [\![0, K_1 - 1]\!], \dots, k_{d,o} \in [\![0, K_d - 1]\!]$, such that $\boldsymbol{x} \in \times_{j\in[\![d]\!]}[1 + k_{j,o}/K_j, 1 + (k_{j,o}+1)/K_j]$ and hence for $\boldsymbol{k}_o := (k_{1,o}, \dots, k_{d,o})^T$

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = f_o(\boldsymbol{x}) + \delta \mathrm{F}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}}^{-1/2} \theta_{\boldsymbol{k}_o+\boldsymbol{1}} \prod_{j\in[\![d]\!]} \psi_{k_{o,j},K_j,\gamma_j}(x_j) \geq \exp(-2d) - \delta \prod_{j\in[\![d]\!]} 2^{\gamma_j} C_{\gamma_j,\infty} c_{\gamma_j}$$

since $\|\psi_{k_j,K_j,\gamma_j}\|_\infty \leq 2^{\gamma_j} C_{\gamma_j,\infty} c_{\gamma_j} K_j^{\gamma_j}$ for any $k_j \in [\![0, K_j - 1]\!]$ and $j \in [\![d]\!]$ and $\mathrm{F}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}} \geq 1$. Choosing $\delta \leq \delta_o(\psi, \boldsymbol{\gamma}) = \exp(-2d)/(\prod_{j\in[\![d]\!]} 2^{\gamma_j} C_{\gamma_j,\infty} c_{\gamma_j})$ ensures $f_{\boldsymbol{\theta}}(x) \geq 0$ for all $x \in \mathbb{R}_+$. □

Proof of Lemma 3.5.12. Our proof starts with the observation that $f_o(\boldsymbol{x}) = \prod_{j\in[\![d]\!]} f_{o,j}(x_j)$ where $f_{o,j}(x_j) := \exp(-x_j)$ if $j \in \mathcal{I}$ and $f_{o,j}(x_j) := x_j \exp(-x_j)$ else, for all $\boldsymbol{x} \in \mathbb{R}_+^d$. By the definition of the multivariate Mellin transform, compare (2.3), we see that $f_o \in \mathbb{L}^2(\mathbb{R}_+, \boldsymbol{x}^{\underline{2\boldsymbol{c}-\boldsymbol{1}}}) \cap \mathbb{L}^1(\mathbb{R}_+^d, \boldsymbol{x}^{\underline{\boldsymbol{c}-\boldsymbol{1}}})$ holds for every $\boldsymbol{c} \in \mathbb{R}_+^d$, and that for all $t \in \mathbb{R}^d$ we have

$$\mathcal{M}_{\boldsymbol{c}}[f_o](\boldsymbol{t}) = \prod_{j\in[\![d]\!]} \mathcal{M}_{c_j}[f_{o,j}](t_j) = \prod_{j\in[\![d]\!]} \Gamma(c_j + it_j) \cdot \prod_{j\in\mathcal{I}^c} (c_j + it_j).$$

Now by applying the Stirling formula (see also Belomestny and Goldenshluger (2020)) we get $|\Gamma(c_j + it_j)| \sim |t_j|^{c_j+1/2} \exp(-\pi|t_j|/2)$, $|t| \geq 2$. Thus for every $\boldsymbol{s} \in \mathbb{N}^d$ there exists $L_{\boldsymbol{s},\boldsymbol{c}}$ such that $|f_o|_{\boldsymbol{s}}^2 \leq L$ for all $L \geq L_{\boldsymbol{s},\boldsymbol{c}}$.
Next, we consider $|f_o - f_{\boldsymbol{\theta}}|_{\boldsymbol{s}}$. Again we see that, $f_o - f_{\boldsymbol{\theta}} \in C_0^\infty(\mathbb{R}_+^d) \subset \mathbb{L}^2(\mathbb{R}_+^d, \boldsymbol{x}^{\underline{2\boldsymbol{c}-\boldsymbol{1}}}) \cap \mathbb{L}^1(\mathbb{R}_+^d, \boldsymbol{x}^{\underline{\boldsymbol{c}-\boldsymbol{1}}})$ with

$$\mathcal{M}_{\boldsymbol{c}}[f_o - f_{\boldsymbol{\theta}}](\boldsymbol{t}) = \delta \mathrm{F}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}}^{-1/2} \sum_{\boldsymbol{k}\in\mathcal{K}} \theta_{\boldsymbol{k}+\boldsymbol{1}} \prod_{j\in[\![d]\!]} \mathcal{M}_{c_j}[\psi_{k_j,K_j,\gamma_j}](t_j), \qquad \text{for all } \boldsymbol{t} \in \mathbb{R}^d.$$

For any fixed $\iota \in [\![d]\!]$, we derive from $\psi_{k_\iota,K_\iota,\gamma_\iota} = \mathrm{S}^{\gamma_\iota}[\psi_{k_\iota,K_\iota}]$ for any $t_\iota \in \mathbb{R}$ that $\mathcal{M}_{c_\iota}[\psi_{K_\iota,k_\iota,\gamma_\iota}](t_\iota) = (c_\iota + it_\iota)^{-s_\iota}\mathcal{M}_{c_\iota}[\psi_{K_\iota,k_\iota,\gamma_\iota+s_\iota}](t_\iota)$. This implies that

$$
\begin{aligned}
&|(1+t_\iota)^{s_\iota}\mathcal{M}_{\boldsymbol{c}}[f_o - f_{\boldsymbol{\theta}}](\boldsymbol{t})|^2 \\
&\qquad \le C_{c_\iota,s_\iota}\delta^2 \mathrm{F}^{-1}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}} \left| \sum_{\boldsymbol{k}\in\mathcal{K}} \theta_{\boldsymbol{k+1}} \mathcal{M}_{c_\iota}[\psi_{k_\iota,K_\iota,\gamma_\iota+s_\iota}](t_\iota) \prod_{j\in[\![d]\!],j\neq\iota} \mathcal{M}_{c_j}[\psi_{k_j,K_j,\gamma_j}](t_j) \right|^2 .
\end{aligned}
$$

Now using that the inverse Mellin operator is linear and by a factorization argument we get

$$
\begin{aligned}
\mathcal{M}^{\dagger}_{\boldsymbol{c}}[\sum_{\boldsymbol{k}\in\mathcal{K}} \theta_{\boldsymbol{k+1}} \mathcal{M}_{c_\iota}[\psi_{k_\iota,K_\iota,\gamma_\iota+s_\iota}](t_\iota) &\prod_{j\in[\![d]\!],j\neq\iota} \mathcal{M}_{c_j}[\psi_{k_j,K_j,\gamma_j}](t_j)](\boldsymbol{x}) \\
&= \sum_{\boldsymbol{k}\in\mathcal{K}} \theta_{\boldsymbol{k+1}} \psi_{k_\iota,K_\iota,\gamma_\iota+s_\iota}(x_\iota) \prod_{j\in[\![d]\!],j\neq\iota} \psi_{k_j,K_j,\gamma_j}(x_j),
\end{aligned}
$$

which implies that due to the disjoint supports of ψ_{k_j,K_j,γ_j} and another factorization argument,

$$
\begin{aligned}
\|(1+t_\iota)^{s_\iota}\mathcal{M}_{\boldsymbol{c}}[f_o - f_{\boldsymbol{\theta}}]\|^2_{\mathbb{R}^d} &\le C_{\boldsymbol{c}_\iota,s_\iota}\delta^2 \mathrm{F}^{-1}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}} \sum_{\boldsymbol{k}\in\mathcal{K}} \|\psi_{k_\iota,K_\iota,\gamma_\iota+s_\iota}\|^2_{x^{2c_j-1}} \prod_{j\in[\![d]\!],j\neq\iota} \|\psi_{k_j,K_j,\gamma_j}\|^2_{x^{2c_j-1}} \\
&\le C_{\boldsymbol{c},\boldsymbol{s},\boldsymbol{\gamma}}\delta^2 \frac{K_\iota^{2s_\iota}}{\sum_{j\in[\![d]\!]} K_j^{2s_j}}
\end{aligned}
$$

since $\|\psi_{k_j,K_j,p}\|^2_{x^{2c_j-1}} = \int_{\mathbb{R}_+} \psi^2_{k_j,K_j,p}(x) x^{2c_j-1} d\lambda(x) \le C_{\gamma_j,s_j,c_j} K_j^{2p-1}$ for any $p \in \mathbb{N}$. We follow $|f_o - f_{\boldsymbol{\theta}}|^2_{\boldsymbol{s},\boldsymbol{c}} \le C_{\boldsymbol{c},\boldsymbol{s},\boldsymbol{\gamma}}$. Finally, we have that $|f_{\boldsymbol{\theta}}|^2_{\boldsymbol{s}} \le 2(|f_o - f_{\boldsymbol{\theta}}|^2_{\boldsymbol{s}} + |f_o|^2_{\boldsymbol{s}}) \le 2(C_{(\boldsymbol{s},\boldsymbol{\gamma},\boldsymbol{c},\delta,\psi)} + L_{\boldsymbol{s}}) =: L_{\boldsymbol{s},\boldsymbol{\gamma},\boldsymbol{c},\delta,1}$.
We consider the moment condition $\mathbb{E}_{f_{\boldsymbol{\theta}}}(\boldsymbol{X}^{\underline{\boldsymbol{2c-2}}}) \le L$. In fact we have

$$
\begin{aligned}
\int_{\mathbb{R}^d_+} \boldsymbol{x}^{\underline{\boldsymbol{2c-2}}} f_{\boldsymbol{\theta}}(\boldsymbol{x}) d\lambda^d(\boldsymbol{x}) &= \prod_{j\in[\![d]\!]} \int_{\mathbb{R}_+} x^{2c_j-2} f_{o,j}(x_j) d\lambda(x_j) \\
&\quad + \delta \mathrm{F}^{-1/2}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}} \sum_{\boldsymbol{k}=(k_1,\dots,k_d)^T\in\mathcal{K}} \theta_{\boldsymbol{k+1}} \prod_{j\in[\![d]\!]} \int_{\mathbb{R}_+} x^{2c_j-2} \psi_{k_j,K_j,\gamma_j}(x_j) d\lambda(x_j) \\
&\le C_{\boldsymbol{c}} + \delta \mathrm{F}^{-1/2}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}} \sum_{\boldsymbol{k}=(k_1,\dots,k_d)^T\in\mathcal{K}} \prod_{j\in[\![d]\!]} C_{\boldsymbol{\gamma},\boldsymbol{c}} K_j^{\gamma-1} \\
&\le C_{\boldsymbol{c}} + \delta C_{\boldsymbol{\gamma},\boldsymbol{c}} =: L_{\boldsymbol{\gamma},\boldsymbol{c},\delta,2}.
\end{aligned}
$$

Now we choose $L_{\boldsymbol{s},\boldsymbol{\gamma},\boldsymbol{c},\delta} := \max(L_{\boldsymbol{s},\boldsymbol{\gamma},\boldsymbol{c},\delta,1}, L_{\boldsymbol{\gamma},\boldsymbol{c},\delta,2})$. □

Proof of Lemma 3.5.13.
Proof of (i): Using that the functions $(\psi_{k_j,K_j,\gamma_j})$ with different index k_j have disjoint supports and a factorization argument we get

$$
\begin{aligned}
\|f_{\boldsymbol{\theta}} - f_{\boldsymbol{\theta}'}\|^2_{\boldsymbol{x}^{2\boldsymbol{c}-\boldsymbol{1}}} &= \delta^2 \mathrm{F}^{-1}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}} \| \sum_{\boldsymbol{k}\in\mathcal{K}} (\theta_{\boldsymbol{k+1}} - \theta'_{\boldsymbol{k+1}}) \prod_{j\in[\![d]\!]} \psi_{k_j,K_j,\gamma_j}(x_j) \|^2_{\boldsymbol{x}^{\underline{\boldsymbol{2c-1}}}} \\
&= \delta^2 F^{-1}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}} \sum_{\boldsymbol{k}\in\mathcal{K}} (\theta_{\boldsymbol{k+1}} - \theta'_{\boldsymbol{k+1}})^2 \prod_{j\in[\![d]\!]} \|\psi_{k_j,K_j,\gamma_j}\|^2_{x^{2c_j-1}} \\
&\ge \delta^2 F^{-1}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}} \rho(\theta_{\boldsymbol{k+1}}, \theta'_{\boldsymbol{k+1}})^2 C_{\boldsymbol{\gamma},\boldsymbol{c}} \boldsymbol{K}^{\underline{\boldsymbol{2\gamma-1}}},
\end{aligned}
$$

where the last step follows if we can show that there exists a $c_{c_j}\mathbb{R}_+$, such that

$$\|\psi_{k_j,K_j,\gamma_j}\|^2_{x^{2c_j-1}} = \int_{\mathbb{R}_+} \psi_{k_j,K_j,\gamma_j}(x)^2 x^{2c_j-1} d\lambda(x) \geq \frac{c_{c_j} K_j^{2\gamma_j-1} \|\psi^{(\gamma_j)}\|^2_{x^0}}{2} \tag{3.36}$$

for K_j big enough. Here $\rho(\boldsymbol{\theta},\boldsymbol{\theta}') := \sum_{\boldsymbol{k}\in\mathcal{K}} \mathbb{1}_{\{\boldsymbol{\theta}_{\boldsymbol{k}+1}\neq\boldsymbol{\theta}'_{\boldsymbol{k}+1}\}}$ denotes the *Hamming distance*.
To show Equation (3.36) we observe that

$$\|\psi_{k_j,K_j,\gamma_j}\|^2_{x^{2c_j+1}} = \sum_{i,\iota\in[\![1,\gamma_j]\!]} c_{i,\gamma_j} c_{\iota,\gamma_j} \int_{\mathbb{R}_+} x^{\iota+i+1} \psi^{(\iota)}_{k_j,K_j}(x)\psi^{(i)}_{k_j,K_j}(x) d\lambda(x),$$

and by defining $\Sigma := \|\psi_{k_j,K_j,\gamma_j}\|^2_{x^{2c_j+1}} - \int_{\mathbb{R}_+}(x^{\gamma_j}\psi^{(\gamma_j)}_{k_j,K_j}(x))^2 x d\lambda(x)$ we can show

$$\begin{aligned}\|\psi_{k_j,K_j,\gamma_j}\|^2_{x^{2c_j-1}} &= \Sigma + \int_{\mathbb{R}_+} (x^{\gamma}\psi^{(\gamma)}_{0,K}(x))^2 x^{2c_j-1} d\lambda(x) \\ &\geq \Sigma + c_{c_j} K^{2\gamma_j-1}\|\psi^{(\gamma_j)}\|^2 \geq \frac{c_{c_j} K_j^{2\gamma_j-1}\|\psi^{(\gamma_j)}\|^2_{x^0}}{2}\end{aligned} \tag{3.37}$$

as soon as $|\Sigma| \leq \frac{c_{c_j} K^{2\gamma-1}\|\psi^{(\gamma_j)}\|^2_{x^0}}{2}$. This is obviously true as soon as $K_j \geq K_o(\gamma_j, c_j, \psi)$ and thus $\|f_{\boldsymbol{\theta}} - f_{\boldsymbol{\theta}'}\|^2_{\boldsymbol{x}^{2\boldsymbol{c}-\boldsymbol{1}}} \geq \delta^2 C_{\boldsymbol{\gamma},\boldsymbol{c}} \boldsymbol{K}^{-\boldsymbol{1}} (\sum_{j\in[\![d]\!]} K_j^{-2s_j})^{-1}\rho(\boldsymbol{\theta},\boldsymbol{\theta}')$ for $\boldsymbol{K} \geq \boldsymbol{K}_o(\psi,\boldsymbol{\gamma},\boldsymbol{c})$, which is understood in a component wise sense.
Now let us interpretate the objects $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ as vectors using the canonical bijection $\mathrm{T}: \boldsymbol{\Theta} \to \{0,1\}^{\prod_{j\in[\![d]\!]} K_j}$. Then $\rho(\boldsymbol{\theta}_{\boldsymbol{k}+\boldsymbol{1}}, \boldsymbol{\theta}'_{\boldsymbol{k}+\boldsymbol{1}})) = \tilde{\rho}(T(\boldsymbol{\theta}_{\boldsymbol{k}+\boldsymbol{1}}), T(\boldsymbol{\theta}'_{\boldsymbol{k}+\boldsymbol{1}}))$, where for $\boldsymbol{\vartheta},\boldsymbol{\vartheta}' \in \{0,1\}^{\prod_{j\in[\![d]\!]} K_j}$, $\tilde{\rho}(\boldsymbol{\vartheta},\boldsymbol{\vartheta}') := \sum_{k=1}^{\prod_{j\in[\![d]\!]} K_j} \mathbb{1}_{\{\vartheta_j\neq\vartheta'_j\}}$. Using the Varshamov-Gilbert Lemma (see Tsybakov (2009)), which states that for $\prod_{j\in[\![d]\!]} K_j \geq 8$ there exists a subset $\{\boldsymbol{\vartheta}^{(0)},\dots,\boldsymbol{\vartheta}^{(M)}\}$ of $\{0,1\}^{\prod_{j\in[\![d]\!]} K_j}$ with $\boldsymbol{\vartheta}^{(0)} = (0,\dots,0)$, such that $\tilde{\rho}(\boldsymbol{\vartheta}^{(j)},\boldsymbol{\vartheta}^{(k)}) \geq \prod_{j\in[\![d]\!]} K_j/8$ for all $j,k \in [\![0,M]\!]$, $j\neq k$ and $M \geq 2^{8^{-1}\prod_{j\in[\![d]\!]} K_j}$. Defining $\boldsymbol{\theta}^{(j)} := T^{-1}(\boldsymbol{\vartheta}^{(j)})$ for $j \in [\![0,M]\!]$ leads to $\|f_{\boldsymbol{\theta}^{(j)}} - f_{\boldsymbol{\theta}^{(l)}}\|^2_{\boldsymbol{x}^{2\boldsymbol{c}-\boldsymbol{1}}} \geq \frac{C_{\boldsymbol{\gamma},\boldsymbol{c}}\delta^2}{\sum_{j\in[\![d]\!]} K_j^{2s_j}}$.
<u>Proof of (ii)</u>: For the second part we have $f_o = f_{\boldsymbol{\theta}^{(0)}}$, and by using $\mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}}, \widetilde{f}_o) \leq \chi^2(\widetilde{f}_{\boldsymbol{\theta}}, \widetilde{f}_o) := \int_{\mathbb{R}^d_+}(\widetilde{f}_{\boldsymbol{\theta}}(x) - \widetilde{f}_o(\boldsymbol{x}))^2/\widetilde{f}_o(\boldsymbol{x}) d\lambda^d(\boldsymbol{x})$ it is sufficient to bound the χ-squared divergence. We notice that since $U_1,\dots,U_n$ are independent we can write $g(\boldsymbol{x}) = \prod_{j\in[\![d]\!]} g_j(x_j)$ for $\boldsymbol{x} \in \mathbb{R}^d_+$. Further, $\widetilde{f}_{\boldsymbol{\theta}} - \widetilde{f}_o$ has support in $[0,2]^d$ since $f_{\boldsymbol{\theta}} - f_o$ has support in $[1,2]^d$ and g has support in $[0,1]^d$. In fact for $\boldsymbol{y} \in \mathbb{R}^d_+$ with $y_j > 2$ for $j \in [\![d]\!]$,

$$\begin{aligned}\widetilde{f}_{\boldsymbol{\theta}}(\boldsymbol{y}) - \widetilde{f}_o(\boldsymbol{y}) &= \int_{\mathbb{R}^d_+} (f_{\boldsymbol{\theta}} - f_o)(\boldsymbol{x})\boldsymbol{x}^{-\boldsymbol{1}} g(\boldsymbol{y}/\boldsymbol{x}) d\lambda^d(\boldsymbol{x}) \\ &= \delta \mathrm{F}^{-1/2}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}} \sum_{\boldsymbol{k}\in\mathcal{K}} \boldsymbol{\theta}_{\boldsymbol{k}+\boldsymbol{1}} \prod_{j\in[\![d]\!]} \int_{y_j}^{\infty} \psi_{k_j,K_j,\gamma_j}(x_j) g_j(y_j/x_j) x_j^{-1} d\lambda(x_j) = 0.\end{aligned}$$

Next we have for any $\boldsymbol{t} \in \mathbb{R}^d$ by application of assumption of Theorem 3.5.8, the convolution theorem and the fact that $\mathcal{M}_{\tilde{c}_j}[\psi_{k_j,K_j,\gamma_j}](t_j) = (\tilde{c}_j + it_j)^{\gamma_j}\mathcal{M}_{\tilde{c}_j}[\psi_{k_j,K_j,0}](t_j)$

$$\begin{aligned}|\mathcal{M}_{\tilde{\boldsymbol{c}}}[\widetilde{f}_{\boldsymbol{\theta}} - \widetilde{f}_o](\boldsymbol{t})| &= |\delta \mathrm{F}^{-1/2}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}} \sum_{\boldsymbol{k}\in\mathcal{K}} \boldsymbol{\theta}_{\boldsymbol{k}+\boldsymbol{1}} \prod_{j\in[\![d]\!]} \mathcal{M}_{\tilde{c}_j}[\psi_{k_j,K_j,\gamma_j}](t_j)\mathcal{M}_{\tilde{c}_j}[g_j](t_j)| \\ &\leq C_{g,\gamma}\delta \mathrm{F}^{-1/2}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}} |\sum_{\boldsymbol{k}\in\mathcal{K}} \boldsymbol{\theta}_{\boldsymbol{k}+\boldsymbol{1}} \prod_{j\in[\![d]\!]} \mathcal{M}_{\tilde{c}_j}[\psi_{k_j,K_j,0}](t_j)|\end{aligned}$$

for all $\boldsymbol{t} \in \mathbb{R}^d$. Applying the Parseval identity, and using the disjoints supports, the factorization property and Equation (3.35) we get

$$\begin{aligned}\chi^2(\widetilde{f}_{\boldsymbol{\theta}}, \widetilde{f}_o) &\leq C_g \|\widetilde{f}_{\boldsymbol{\theta}} - \widetilde{f}_o\|^2_{\underline{\boldsymbol{x}^{2\tilde{\boldsymbol{c}}-\mathbf{1}}}} \\ &\leq C_{g,\boldsymbol{\gamma}} \delta^2 \mathrm{F}^{-1}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}} \| \sum_{\boldsymbol{k} \in \mathcal{K}} \prod_{j \in [\![d]\!]} \mathcal{M}_{\tilde{c}_j}[\psi_{k_j,K_j,0}](t_j)\|^2_{\mathbb{R}^d} \\ &\leq C_{g,\boldsymbol{\gamma}} \delta^2 \mathrm{F}^{-1}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}} \sum_{\boldsymbol{k} \in \mathcal{K}} \prod_{j \in [\![d]\!]} \|\psi_{k_j,K_j,0}\|^2_{x^{2\tilde{c}_j-1}}.\end{aligned}$$

In fact, using that $\|\psi_{k_j,K_j,0}\|^2_{x^{2\tilde{c}_j-1}} \leq K_j^{-1}\|\psi\|^2_{x^0}$ implies $\chi^2(\widetilde{f}_{\boldsymbol{\theta}}, \widetilde{f}_o) \leq C_g \delta^2 \mathrm{F}^{-1}_{\boldsymbol{K},\boldsymbol{\gamma},\boldsymbol{s}}$. Since $M \geq 2^{\prod_{j\in[\![d]\!]} K_j}$ we can deduce that

$$\mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}^{(j)}}, \widetilde{f}_{\boldsymbol{\theta}^{(0)}}) \leq C_{g,\boldsymbol{\gamma}} \delta^2 \log(M) \boldsymbol{K}^{\underline{2\boldsymbol{\gamma}-\mathbf{1}}} (\sum_{j \in [\![d]\!]} K_j^{2s_j})^{-1}.$$

□

Proof of Theorem 3.5.9. Let $\boldsymbol{k} \in \mathcal{K}_n$. By definition of the estimator, (3.29), we have $[-\boldsymbol{k}', \boldsymbol{k}'] = \mathrm{supp}(\mathcal{M}_{\boldsymbol{c}}[f_{\boldsymbol{k}'}])$, for $\boldsymbol{k}' \in \mathcal{K}_n$ and we can find a $\boldsymbol{K}_n \in (\mathbb{N}^*)^d$ such that for all $\boldsymbol{k}' \in \mathcal{K}_n$ holds $[-\boldsymbol{k}', \boldsymbol{k}'] \subseteq [-\boldsymbol{K}_n, \boldsymbol{K}_n]$. Then we have for any $\boldsymbol{k}' \in \mathcal{K}_n$ that $\|\widehat{f}_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} - \|\widehat{f}_{\boldsymbol{k}'}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} = \|\widehat{f}_{\boldsymbol{K}_n} - \widehat{f}_{\boldsymbol{k}'}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}}$ implying with (3.32)

$$\|\widehat{f}_{\widehat{\boldsymbol{k}}} - \widehat{f}_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} + \widehat{\mathrm{pen}}(\widehat{\boldsymbol{k}}) \leq \|\widehat{f}_{\boldsymbol{k}} - \widehat{f}_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} + \widehat{\mathrm{pen}}(\boldsymbol{k}). \tag{3.38}$$

Now for every $\boldsymbol{k}' \in \mathcal{K}_n$ we have $\|\widehat{f}_{\boldsymbol{k}'} - f_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} = \|\widehat{f}_{\boldsymbol{k}'} - \widehat{f}_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} + \|\widehat{f}_{\boldsymbol{K}_n} - f_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} + 2\langle \widehat{f}_{\boldsymbol{k}'} - \widehat{f}_{\boldsymbol{K}_n}, \widehat{f}_{\boldsymbol{K}_n} - f_{\boldsymbol{K}_n}\rangle_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}}$, which combined with (3.38) implies

$$\|\widehat{f}_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} - \|\widehat{f}_{\boldsymbol{k}} - f_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} \leq \widehat{\mathrm{pen}}(\boldsymbol{k}) - \widehat{\mathrm{pen}}(\widehat{\boldsymbol{k}}) + 2\langle \widehat{f}_{\widehat{\boldsymbol{k}}} - \widehat{f}_{\boldsymbol{k}}, \widehat{f}_{\boldsymbol{K}_n} - f_{\boldsymbol{K}_n}\rangle_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}}. \tag{3.39}$$

Since $\langle \widehat{f}_{\widehat{\boldsymbol{k}}} - \widehat{f}_{\boldsymbol{k}}, \widehat{f}_{\boldsymbol{K}_n} - f_{\boldsymbol{K}_n}\rangle_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} = \|\widehat{f}_{\widehat{\boldsymbol{k}}} - f_{\widehat{\boldsymbol{k}}}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} + \langle f_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{k}}, \widehat{f}_{\boldsymbol{K}_n} - f_{\boldsymbol{K}_n}\rangle_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} - \|\widehat{f}_{\boldsymbol{k}} - f_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}}$ we get

$$\begin{aligned}\|\widehat{f}_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} \leq \|f_{\boldsymbol{k}} - f_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} - \|\widehat{f}_{\boldsymbol{k}} - f_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} + 2\langle f_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{k}}, \widehat{f}_{\boldsymbol{K}_n} - f_{\boldsymbol{K}_n}\rangle_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} \\ + \widehat{\mathrm{pen}}(\boldsymbol{k}) + 2\|\widehat{f}_{\widehat{\boldsymbol{k}}} - f_{\widehat{\boldsymbol{k}}}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} - \widehat{\mathrm{pen}}(\widehat{\boldsymbol{k}}).\end{aligned} \tag{3.40}$$

We now consider the term $|2\langle f_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{k}}, \widehat{f}_{\boldsymbol{K}_n} - f_{\boldsymbol{K}_n}\rangle_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}}|$. First we remind that for any $\boldsymbol{k}' \in \mathcal{K}_n$

$$\|\widehat{f}_{\boldsymbol{k}'} - f_{\boldsymbol{k}'}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathbb{1}_{[-\boldsymbol{k}',\boldsymbol{k}]}(\boldsymbol{t}) \frac{|\mathcal{M}_{\boldsymbol{c}}[f_{\boldsymbol{Y}}](\boldsymbol{t}) - \widehat{\mathcal{M}}_{\boldsymbol{c}}(\boldsymbol{t})|^2}{|\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})|^2} d\lambda^d(\boldsymbol{t}).$$

Setting $Q^* := [-\widehat{\boldsymbol{k}}, \widehat{\boldsymbol{k}}] \cup [-\boldsymbol{k}, \boldsymbol{k}]$ we have $\mathcal{M}_{\boldsymbol{c}}[f_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{k}}] = \mathcal{M}_{\boldsymbol{c}}[f](\mathbb{1}_{[-\widehat{\boldsymbol{k}},\widehat{\boldsymbol{k}}]} - \mathbb{1}_{[-\boldsymbol{k},\boldsymbol{k}]})$ implying that $\mathrm{supp}(\mathcal{M}_{\boldsymbol{c}}[f_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{k}}]) \subseteq Q^* \subseteq [-\boldsymbol{K}_n, \boldsymbol{K}_n]$ by definition of $\boldsymbol{K}_n$. Using that $2ab \leq a^2 + b^2$ we deduce

$$\begin{aligned}|2\langle f_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{k}}, \widehat{f}_{\boldsymbol{K}_n} - f_{\boldsymbol{K}_n}\rangle_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}}| &= \frac{2}{(2\pi)^d} \left| \int_{Q^*} \mathcal{M}_{\boldsymbol{c}}[f_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{k}}](\boldsymbol{t}) \frac{\widehat{\mathcal{M}}_{\boldsymbol{c}}(-\boldsymbol{t}) - \mathcal{M}_{\boldsymbol{c}}[f_{\boldsymbol{Y}}](-\boldsymbol{t})}{\mathcal{M}_{\boldsymbol{c}}[g](-\boldsymbol{t})} d\lambda^d(\boldsymbol{t}) \right| \\ &\leq \frac{1}{4}\|f_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} + \frac{4}{(2\pi)^d} \int_{Q^*} \frac{|\widehat{\mathcal{M}}_{\boldsymbol{c}}(\boldsymbol{t}) - \mathcal{M}_{\boldsymbol{c}}[f_{\boldsymbol{Y}}](\boldsymbol{t})|^2}{|\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})|^2} d\lambda^d(\boldsymbol{t}) \\ &\leq \frac{1}{2}\|f_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} + \frac{1}{2}\|f_{\boldsymbol{k}} - f_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} + 4\|\widehat{f}_{\boldsymbol{k}} - f_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} \\ &+ 4\|\widehat{f}_{\widehat{\boldsymbol{k}}} - f_{\widehat{\boldsymbol{k}}}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}}\end{aligned}$$

using that $\mathbb{1}_{Q^*} \leq \mathbb{1}_{[-\widehat{\boldsymbol{k}},\widehat{\boldsymbol{k}}]} + \mathbb{1}_{[-\boldsymbol{k},\boldsymbol{k}]}$. Since $\|f_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} + \|\widehat{f}_{\widehat{\boldsymbol{k}}} - f_{\widehat{\boldsymbol{k}}}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} = \|\widehat{f}_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2c-1}}}$ we get

$$|2\langle f_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{k}}, \widehat{f}_{\boldsymbol{K}_n} - f_{\boldsymbol{K}_n}\rangle_{\underline{\boldsymbol{x}^{2c-1}}}| \leq \frac{1}{2}\|\widehat{f}_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} + \frac{1}{2}\|f_{\boldsymbol{K}_n} - f_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} \\ + 4\|\widehat{f}_{\boldsymbol{k}} - f_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} + \frac{7}{2}\|\widehat{f}_{\widehat{\boldsymbol{k}}} - f_{\widehat{\boldsymbol{k}}}\|^2_{\underline{\boldsymbol{x}^{2c-1}}},$$

implying that

$$\|\widehat{f}_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} \leq 3\|f_{\boldsymbol{k}} - f_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} + 6\|\widehat{f}_{\boldsymbol{k}} - f_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} + 2\widehat{\mathrm{pen}}(\boldsymbol{k}) \\ + 11\|\widehat{f}_{\widehat{\boldsymbol{k}}} - f_{\widehat{\boldsymbol{k}}}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} - 2\widehat{\mathrm{pen}}(\widehat{\boldsymbol{k}}). \tag{3.41}$$

Now since $\mathbb{E}^n_{f_{\boldsymbol{Y}}}(\widehat{\mathrm{pen}}(\boldsymbol{k})) = \mathrm{pen}(\boldsymbol{k})$ and $6\mathbb{E}^n_{f_{\boldsymbol{Y}}}(\|\widehat{f}_{\boldsymbol{k}} - f_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}^{2c-1}}}) \leq 6\chi^{-1}\mathrm{pen}(\boldsymbol{k}) \leq \mathrm{pen}(\boldsymbol{k})$ we deduce from (3.41) that

$$\mathbb{E}^n_{f_{\boldsymbol{Y}}}(\|\widehat{f}_{\widehat{\boldsymbol{k}}} - f_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2c-1}}}) \leq 3\left(\|f_{\boldsymbol{K}_n} - f_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} + \mathrm{pen}(\boldsymbol{k})\right) + 11\mathbb{E}^n_{f_{\boldsymbol{Y}}}(\left(\|\widehat{f}_{\widehat{\boldsymbol{k}}} - f_{\widehat{\boldsymbol{k}}}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} - \frac{1}{6}\widehat{\mathrm{pen}}(\widehat{\boldsymbol{k}})\right)_+) \\ \leq 3\left(\|f_{\boldsymbol{K}_n} - f_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} + \mathrm{pen}(\boldsymbol{k})\right) + 11\mathbb{E}^n_{f_{\boldsymbol{Y}}}(\left(\|\widehat{f}_{\widehat{\boldsymbol{k}}} - f_{\widehat{\boldsymbol{k}}}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} - \frac{1}{12}\mathrm{pen}(\widehat{\boldsymbol{k}})\right)_+) \\ + \mathbb{E}^n_{f_{\boldsymbol{Y}}}((\mathrm{pen}(\widehat{\boldsymbol{k}}) - 2\widehat{\mathrm{pen}}(\widehat{\boldsymbol{k}}))_+).$$

The two expectations on the right hand side of the last inequality can be bounded using the following Lemma.

Lemma 3.5.14:
Let the assumptions of Theorem 3.5.9 hold true. Then

(i) $\mathbb{E}^n_{f_{\boldsymbol{Y}}}(\left(\|\widehat{f}_{\widehat{\boldsymbol{k}}} - f_{\widehat{\boldsymbol{k}}}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} - \frac{1}{12}\mathrm{pen}(\widehat{\boldsymbol{k}})\right)_+) \leq C_{f,g,\sigma,\mathbb{E}_{f_Y}(\boldsymbol{Y}^{\underline{7(c-1)}})} n^{-1}$,

(ii) $\mathbb{E}^n_{f_{\boldsymbol{Y}}}((\mathrm{pen}(\widehat{\boldsymbol{k}}) - 2\widehat{\mathrm{pen}}(\widehat{\boldsymbol{k}}))_+) \leq C_{\sigma,\mathbb{E}_{f_Y}(\boldsymbol{Y}^{\underline{4(c-1)}})} n^{-1}$.

Consequently, we have

$$\mathbb{E}^n_{f_{\boldsymbol{Y}}}(\|\widehat{f}_{\widehat{\boldsymbol{k}}} - f\|_{\underline{\boldsymbol{x}^{2c-1}}})^2) \leq \|f - f_{\boldsymbol{K}_n}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} + 3\left(\|f_{\boldsymbol{K}_n} - f_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} + \mathrm{pen}(\boldsymbol{k})\right) + \frac{C_{f,g}}{n} \\ \leq 3\left(\|f - f_{\boldsymbol{k}}\|^2_{\underline{\boldsymbol{x}^{2c-1}}} + \mathrm{pen}(\boldsymbol{k})\right) + \frac{C_{f,g}}{n}.$$

Taking now the infimum over all $\boldsymbol{k} \in \mathcal{K}_n$ implies the claim. □

Proof of Lemma 3.5.14. We start by proving (i). Let us therefore define the set $\mathbb{U} := \{h \in \mathbb{L}^2(\mathbb{R}^d_+, \underline{\boldsymbol{x}^{2c-1}}) : \|h\|_{\underline{\boldsymbol{x}^{2c-1}}} \leq 1\}$. Then for $\boldsymbol{k}' \in \mathbb{R}^d_+$, $\|\widehat{f}_{\boldsymbol{k}'} - f_{\boldsymbol{k}'}\|_{\underline{\boldsymbol{x}^{2c-1}}} = \sup_{h\in\mathbb{U}}\langle \widehat{f}_{\boldsymbol{k}'} - f_{\boldsymbol{k}'}, h\rangle_{\underline{\boldsymbol{x}^{2c-1}}}$, where

$$\langle \widehat{f}_{\boldsymbol{k}'} - f_{\boldsymbol{k}'}, h\rangle_{\underline{\boldsymbol{x}^{2c-1}}} = (2\pi)^{-d}\int_{[-\boldsymbol{k}',\boldsymbol{k}']}(\widehat{\mathcal{M}}_{\boldsymbol{c}}(\boldsymbol{t}) - \mathbb{E}^n_{f_Y}(\widehat{\mathcal{M}}_{\boldsymbol{c}}(\boldsymbol{t}))\frac{\mathcal{M}_{\boldsymbol{c}}[h](-\boldsymbol{t})}{\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})}d\lambda^d(\boldsymbol{t})$$

by application of the Plancherel identity. Now for a sequence $(d_n)_{n\in\mathbb{N}}$ we decompose the estimator $\widehat{\mathcal{M}}_{\boldsymbol{c}}(\boldsymbol{t})$ into

$$\begin{aligned}\widehat{\mathcal{M}}_{\boldsymbol{c}}(\boldsymbol{t}) :&= n^{-1}\sum_{j\in[\![n]\!]}\boldsymbol{Y}_j^{\underline{\boldsymbol{c}-\mathbf{1}+i\boldsymbol{t}}}\mathbb{1}_{(0,d_n)}(\boldsymbol{Y}_j^{\underline{\boldsymbol{c}-\mathbf{1}}}) + n^{-1}\sum_{j\in[\![n]\!]}\boldsymbol{Y}_j^{\underline{\boldsymbol{c}-\mathbf{1}+i\boldsymbol{t}}}\mathbb{1}_{[d_n,\infty)}(\boldsymbol{Y}_j^{\underline{\boldsymbol{c}-\mathbf{1}}})\\ &=: \widehat{\mathcal{M}}_{\boldsymbol{c},1}(\boldsymbol{t}) + \widehat{\mathcal{M}}_{\boldsymbol{c},2}(\boldsymbol{t}).\end{aligned}$$

Setting

$$\nu_{\boldsymbol{k}',i}(h) := \frac{1}{(2\pi)^d}\int_{[-\boldsymbol{k}',\boldsymbol{k}']}(\widehat{\mathcal{M}}_{\boldsymbol{c},i}(\boldsymbol{t}) - \mathbb{E}^n_{f_{\boldsymbol{Y}}}(\widehat{\mathcal{M}}_{\boldsymbol{c},i}(\boldsymbol{t}))\frac{\mathcal{M}_{\boldsymbol{c}}[h](-\boldsymbol{t})}{\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})}d\lambda^d(\boldsymbol{t}), \quad h\in\mathbb{U}, i\in\{1,2\},$$

we can deduce that

$$\begin{aligned}\mathbb{E}^n_{f_{\boldsymbol{Y}}}\left(\left(\|\widehat{f}_{\widehat{\boldsymbol{k}}} - f_{\widehat{\boldsymbol{k}}}\|^2_{\underline{\boldsymbol{x}^{2\boldsymbol{c}-\mathbf{1}}}} - \frac{1}{12}\mathrm{pen}(\widehat{\boldsymbol{k}})\right)_+\right) &\leq 2\mathbb{E}^n_{f_{\boldsymbol{Y}}}\left(\left(\sup_{h\in\mathbb{U}}\nu_{\widehat{\boldsymbol{k}},1}(h)^2 - \frac{1}{24}\mathrm{pen}(\widehat{\boldsymbol{k}})\right)_+\right)\\ &\quad + 2\mathbb{E}^n_{f_{\boldsymbol{Y}}}(\sup_{h\in\mathbb{U}}\nu_{\widehat{\boldsymbol{k}},2}(h)^2)).\end{aligned}\tag{3.42}$$

We start by bounding the first summand. To do so, we see that

$$\mathbb{E}^n_{f_{\boldsymbol{Y}}}\left(\left(\sup_{h\in\mathbb{U}}\nu_{\widehat{\boldsymbol{k}},1}(h)^2 - \frac{1}{24}\mathrm{pen}(\widehat{\boldsymbol{k}})\right)_+\right) \leq \sum_{\boldsymbol{k}'\in\mathcal{K}_n}\mathbb{E}^n_{f_{\boldsymbol{Y}}}\left(\left(\sup_{h\in\mathbb{U}}\nu_{\boldsymbol{k}',1}(h)^2 - \frac{1}{24}\mathrm{pen}(\boldsymbol{k}')\right)_+\right).$$

To control each summand we apply the Talagrand inequality, see Remark 3.1.14, which can be done since there exists a dense subset of $\mathbb{U}$. For each $\boldsymbol{k}'\in\mathbb{R}^d_+$ and $h\in\mathbb{U}$ we set

$$\nu_h(\boldsymbol{y}) := \frac{1}{(2\pi)^d}\int_{[-\boldsymbol{k}',\boldsymbol{k}']}\boldsymbol{y}^{\underline{\boldsymbol{c}-\mathbf{1}+i\boldsymbol{t}}}\mathbb{1}_{(0,d_n)}(\boldsymbol{y}^{\underline{\boldsymbol{c}-\mathbf{1}}})\frac{\mathcal{M}_{\boldsymbol{c}}[h](-\boldsymbol{t})}{\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})}d\lambda^d(\boldsymbol{t}), \quad \boldsymbol{y}\in\mathbb{R}^d_+.$$

So, $\overline{\nu}_h = \nu_{\boldsymbol{k}',1}(h)$ in the notation of Remark 3.1.14. Thus we need to determine the parameters τ, Ψ^2, ψ. Let us begin with Ψ. For $h\in\mathbb{U}$ we have $1\geq\|h\|^2_{\boldsymbol{x}^{2\boldsymbol{c}-1}} = (2\pi)^{-d}\|\mathcal{M}_c[h]\|_{\mathbb{R}^d}$. Using the Cauchy-Schwarz inequality delivers

$$\mathbb{E}^n_{f_{\boldsymbol{Y}}}(\sup_{h\in\mathbb{U}}\overline{\nu}_h^2) \leq (2\pi)^{-d}\int_{[-\boldsymbol{k}',\boldsymbol{k}']}\frac{\mathbb{E}^n_{f_{\boldsymbol{Y}}}(|\widehat{\mathcal{M}}_{\boldsymbol{c},1}(\boldsymbol{t}) - \mathbb{E}^n_{f_Y}(\widehat{\mathcal{M}}_{\boldsymbol{c},1}(\boldsymbol{t}))|^2)}{|\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})|^2}d\lambda^d(\boldsymbol{t}) \leq \sigma n^{-1}\Delta_g(\boldsymbol{k}') =: \Psi^2.$$

Now for τ we see that $\mathbb{V}\mathrm{ar}^n_{f_{\boldsymbol{Y}}}(\nu_h(\boldsymbol{Y}_1)) \leq \mathbb{E}^n_{f_{\boldsymbol{Y}}}(\nu_h^2(\boldsymbol{Y}_1)) \leq \|f_Y\boldsymbol{x}^{\underline{2\boldsymbol{c}-\mathbf{1}}}\|_\infty\|\nu_h\|^2_{\underline{\boldsymbol{x}^{\mathbf{1}-2\boldsymbol{c}}}}$. Further,

$$\|\nu_h\|^2_{\underline{\boldsymbol{x}^{\mathbf{1}-2\boldsymbol{c}}}} = (2\pi)^{-d}\int_{[-\boldsymbol{k}',\boldsymbol{k}']}|\mathcal{M}_{\boldsymbol{c}}[h](\boldsymbol{t})|^2|\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})|^{-2}d\boldsymbol{t} \leq \|\mathbb{1}_{[-\boldsymbol{k}',\boldsymbol{k}']}\mathcal{M}_{\boldsymbol{c}}[g]^{-2}\|_\infty.$$

Thus we choose $\tau := \|f_Y\boldsymbol{x}^{\underline{2\boldsymbol{c}-\mathbf{1}}}\|_\infty\|\mathbb{1}_{[-\boldsymbol{k}',\boldsymbol{k}']}\mathcal{M}_{\boldsymbol{c}}[g]^{-2}\|_\infty$. Let us now consider ψ^2. We have for any $\boldsymbol{y}\in\mathbb{R}^d_+$,

$$|\nu_h(\boldsymbol{y})|^2 = (2\pi)^{-2d}\left|\int_{[-\boldsymbol{k}',\boldsymbol{k}']}\boldsymbol{y}^{\underline{\boldsymbol{c}-\mathbf{1}+i\boldsymbol{t}}}\mathbb{1}_{(0,d_n)}(\boldsymbol{y}^{\underline{\boldsymbol{c}-\mathbf{1}}})\frac{\mathcal{M}_{\boldsymbol{c}}[h](-\boldsymbol{t})}{\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})}d\boldsymbol{t}\right|^2 \leq d_n^2\Delta_g(\boldsymbol{k}) =: \psi^2,$$

since $\|h\|_{\underline{x^{2c-1}}} \leq 1$ and $|\underline{y^{it}}| = 1$. Applying now the Talagrand inequality we get

$$\begin{aligned}\mathbb{E}^n_{f_{\boldsymbol{Y}}}((\sup_{h\in\mathbb{U}} \overline{\nu}_h^2 - 6\Psi^2)_+) &\leq \frac{C_{f_{\boldsymbol{Y}}}}{n}\left(\underline{\boldsymbol{k}^{2\gamma}} \exp(-C_{f_{\boldsymbol{Y}},\sigma}\underline{\boldsymbol{k}^{\mathbf{1}}}) + d_n^2 \exp(-\frac{\sqrt{n\sigma}}{100 d_n})\right) \\ &\leq \frac{C_{f_{\boldsymbol{Y}},\sigma}}{n}\left(\underline{\boldsymbol{k}^{2\gamma}} \exp(-C_{f_{\boldsymbol{Y}},\sigma}\underline{\boldsymbol{k}^{\mathbf{1}}}) + n^{-d}\right)\end{aligned}$$

for the choice $d_n := \sqrt{n\sigma}/(100\log(n^{d+1}))$. For $\chi \geq 144$ we can conclude that

$$\begin{aligned}\mathbb{E}^n_{f_{\boldsymbol{Y}}}(\left(\sup_{h\in\mathbb{U}} \nu_{\widehat{\boldsymbol{k}},1}(h)^2 - \frac{\chi}{24}\sigma\Delta_g(\widehat{\boldsymbol{k}})n^{-1}\right)_+) &\leq \sum_{\boldsymbol{k}\in\mathcal{K}_n} \frac{C_{f_{\boldsymbol{Y}},\sigma}}{n}\left(\underline{\boldsymbol{k}^{2\gamma}} \exp(-C_{f_{\boldsymbol{Y}},\sigma}\underline{\boldsymbol{k}^{\mathbf{1}}}) + \frac{1}{n^d}\right) \\ &\leq \frac{C_{f_{\boldsymbol{Y}},\sigma,\gamma}}{n},\end{aligned}$$

since $|\mathcal{K}_n| \leq n^d$. For the second summand in (3.28) we get for any $\boldsymbol{k}' \in \mathcal{K}_n$ and $h \in \mathbb{U}$,

$$\begin{aligned}|\nu_{\widehat{\boldsymbol{k}},2}(h)|^2 &\leq (2\pi)^{-d}\int_{[-\widehat{\boldsymbol{k}},\widehat{\boldsymbol{k}}]} |\widehat{\mathcal{M}}_{\boldsymbol{c},2}(\boldsymbol{t}) - \mathbb{E}^n_{f_{\boldsymbol{Y}}}(\widehat{\mathcal{M}}_{\boldsymbol{c},2}(\boldsymbol{t}))|^2 |\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})|^{-2} d\lambda^d(\boldsymbol{t}) \\ &\leq \sum_{\boldsymbol{k}'\in\mathcal{K}_n} (2\pi)^{-d}\int_{[-\boldsymbol{k}',\boldsymbol{k}']} |\widehat{\mathcal{M}}_{\boldsymbol{c},2}(\boldsymbol{t}) - \mathbb{E}^n_{f_{\boldsymbol{Y}}}(\widehat{\mathcal{M}}_{\boldsymbol{c},2}(\boldsymbol{t}))|^2 |\mathcal{M}_{\boldsymbol{c}}[g](\boldsymbol{t})|^{-2} d\lambda^d(\boldsymbol{t}).\end{aligned}$$

Thus, we have for any $u > 0$

$$\mathbb{E}^n_{f_{\boldsymbol{Y}}}(\sup_{h\in\mathbb{U}} \nu_{\widehat{\boldsymbol{k}},2}(h)^2)) \leq \sum_{\boldsymbol{k}'\in\mathcal{K}_n} \frac{\Delta_g(\boldsymbol{k}')}{n}\mathbb{E}_{f_{\boldsymbol{Y}}}(\underline{\boldsymbol{Y}_1^{2\boldsymbol{c}-\boldsymbol{2}}}\mathbb{1}_{[d_n,\infty)}(\underline{\boldsymbol{Y}_1^{\boldsymbol{c}-\boldsymbol{1}}})) \leq \frac{C_g|\mathcal{K}_n|}{d_n^u}\mathbb{E}_{f_{\boldsymbol{Y}}}(\underline{\boldsymbol{Y}_1^{(2+u)(\boldsymbol{c}-\boldsymbol{1})}}).$$

Now under assumption **[G1]** we have $|\mathcal{K}_n| \leq |\{\boldsymbol{k} \in \mathbb{N}^d : \underline{\boldsymbol{k}^{2\gamma+\mathbf{1}}} \leq c_g n\}| \leq C_g n \log(n)^{d-1}$, compare Dussap (2023). Now choosing $u = 5$ implies

$$\mathbb{E}^n_{f_{\boldsymbol{Y}}}(\sup_{h\in\mathbb{U}} \nu_{\widehat{\boldsymbol{k}},2}(h)^2)) \leq C_{g,\sigma}\mathbb{E}_{f_{\boldsymbol{Y}}}(\underline{\boldsymbol{Y}_1^{7(\boldsymbol{c}-\boldsymbol{1})}})n^{-1}.$$

To finish the proof we still need to show (ii). To do so, we define the event $\Omega := \{|\widehat{\sigma} - \sigma| \leq \sigma/2\}$. On Ω holds $\sigma \leq 2\widehat{\sigma} \leq 3\sigma$ and we deduce

$$(\mathrm{pen}(\widehat{\boldsymbol{k}}) - 2\widehat{\mathrm{pen}}(\widehat{\boldsymbol{k}}))_+ \leq \frac{\Delta_g(\widehat{\boldsymbol{k}})}{n}\chi(\sigma - 2\widehat{\sigma})_+\mathbb{1}_{\Omega^c} \leq \chi(\sigma - 2\widehat{\sigma})_+\mathbb{1}_{\Omega^c} \leq \chi 2|\sigma - \widehat{\sigma}|\mathbb{1}_{\Omega^c},$$

since $\widehat{\boldsymbol{k}} \in \mathcal{K}_n$. Now, since $|\sigma - \widehat{\sigma}|\mathbb{1}_{\Omega c} \leq 2|\sigma - \widehat{\sigma}|^2\sigma^{-1}$ we get

$$\mathbb{E}^n_{f_{\boldsymbol{Y}}}((\mathrm{pen}(\widehat{\boldsymbol{k}}) - 2\widehat{\mathrm{pen}}(\widehat{\boldsymbol{k}}))_+) \leq \frac{4\chi}{\sigma}\mathrm{Var}^n_{f_{\boldsymbol{Y}}}(\widehat{\sigma}) = \frac{4\chi\mathbb{E}_{f_{\boldsymbol{Y}}}(\underline{\boldsymbol{Y}_1^{4(\boldsymbol{c}-\boldsymbol{1})}})}{\sigma n}.$$

□

3.6 Under multiplicative measurement errors for stationary processes

3.6.1 Introduction

Again, we consider the multiplicative measurement error model, that is

$$Y = XU,$$

where the density f of X is our quantity of interested and U is < second positive random variable independent of X with density g. In contrary to Section 3.2 we assume that we have access to a sample $(Y_j)_{j\in[\![n]\!]}$, $Y_j = X_jU_j$, where $(X_j)_{j\in[\![n]\!]}$ is a strictly stationary process with marginal density f and $(U_j)_{j\in[\![n]\!]}$ is an i.i.d. sample of U independent of $(X_j)_{j\in[\![n]\!]}$.
In other words, we will not restrict ourselves to the case of an i.i.d. sample $(X_j)_{j\in[\![n]\!]}$ but some dependency structure on the process $(X_j)_{j\in[\![n]\!]}$ and analyze its influence on the risk of the estimator $(\widehat{f}_k)_{k\in[\![n]\!]}$, (3.14), given by

$$\widehat{f}_k(x) := \frac{1}{2\pi}\int_{[-k,k]} x^{-c-it}\frac{\widehat{\mathcal{M}}_c(t)}{\mathcal{M}_c[g](t)}d\lambda(t), \quad x\in\mathbb{R}_+,$$

under the assumption **[G0]**, introduced in Section 3.2, which implies the well-definedness of our estimator. Before analyzing the risk of our estimator, let us briefly recapitulate the definition of strict stationarity and introduce two notions of dependence, which will be considered througout this section. For a more detailed introduction we refer to Rio (2013).

Definition 3.6.1 (Strict stationary process)**:**
The process $(X_j)_{j\in\mathbb{N}}$ is said to be STRICTLY STATIONARY, if for any $j\in\mathbb{N}$ and $S\subset\mathbb{N}$ finite, it holds

$$\mathbb{P}^{(X_{j+s}:s\in S)} = \mathbb{P}^{(X_s:s\in S)}.$$

Furthermore, we call a process $(X_t)_{t\geq 0}$ STRICTLY STATIONARY, if for any $t\in\mathbb{R}_+$ and $S\subset\mathbb{R}_+$ finite, it holds

$$\mathbb{P}^{(X_{t+s}:s\in S)} = \mathbb{P}^{(X_s:s\in S)}.$$

As a direct consequence, we can see that the marginals distributions $(\mathbb{P}^{X_j})_{j\in\mathbb{N}}$ of the process $(X_j)_{j\in\mathbb{N}}$ are identical. A trivial example of stationary processes are i.i.d. processes considered in Section 3.2.

Definition 3.6.2 (β-mixing coefficient of σ-fields)**:**
For two sigma fields $\mathcal{U}$ and $\mathcal{V}$ over some probability space Ω we define the β-MIXING COEFFICIENT by

$$\beta^{\text{mix}}(\mathcal{U},\mathcal{V}) := \frac{1}{2}\sup_{(U_p)_{p\in P},(V_q)_{q\in Q}}\sum_{(p,q)\in P\times Q}|\mathbb{P}(U_p\cap V_q)-\mathbb{P}(U_p)\mathbb{P}(V_q)|,$$

where the supremum is taking over all finite partitions $(U_p)_{p\in P}$, respectively $(V_q)_{q\in Q}$, of Ω with $(U_p)_{p\in P}\subset\mathcal{U}$, respectively $(V_q)_{q\in Q}\subset\mathcal{V}$.

As a direct consequence of the definition, we see that for two independent σ–fields $\mathcal{U}$ and $\mathcal{V}$ the β–mixing coefficient is 0. On the other hand, if $\beta^{\text{mix}}(\mathcal{U},\mathcal{V}) = 0$ then $\mathcal{U}$ and $\mathcal{V}$ are independent σ–fields. Thus β^{mix} measures dependency between two σ–fields. Now we will first introduce the notion of a β–mixing process $(X_j)_{j\in\mathbb{N}_o}$, $\mathbb{N}_o := \mathbb{N}\cup\{0\}$, and then precise the definition of a β–mixing sample $(X_j)_{j\in[\![n]\!]}$.

Definition 3.6.3 (β–mixing process)**:**
Let now $(X_k)_{k\in\mathbb{N}_o}$ be a strictly stationary sequence of real-valued random variables and let us define the positive real-valued sequence $(\beta(k))_{k\in\mathbb{N}}$ by

$$\beta(k) := \beta(X_0, X_k) := \beta^{\text{mix}}(\sigma(X_0), \sigma(X_k))$$

where $\sigma(X_0)$, respectively $\sigma(X_k)$, is the σ field generated by X_0, respectively X_k. Furthermore, we call a process $(X_k)_{k\in\mathbb{N}_0}$ β–MIXING, if $\lim_{k\to\infty}\beta(k) = 0$ and a finite sample $(X_k)_{k\in[\![n]\!]}$ β–mixing, if its a finite subsequence of a β–mixing process.
In analogy, we define for a strictly stationary process $(X_t)_{t\geq 0}$ the family $(\beta(t))_{t\geq 0}$ through

$$\beta(t) := \beta^{\text{mix}}(\sigma(X_0), \sigma(X_t)).$$

Many linear processes of current interest, such as GARCH or ARMA processes, have absolutely summable $\beta(k)$, cf. Bradley (2005), Fryzlewicz and Subba Rao (2011) or Doukhan (1994).

3.6.2 Estimation strategy

Due to the strict stationarity of $(X_j)_{j\in[\![n]\!]}$ we still have that $\mathbb{E}^n_{f_Y}(\widehat{f}_k) = f_k$ and that the squared bias variance decomposition

$$\mathbb{E}^n_{f_Y}(\|\widehat{f}_k - f\|^2_{x^{2c-1}}) = \|f - f_k\|^2_{x^{2c-1}} + \mathbb{E}^n_{f_Y}(\|\widehat{f}_k - f_k\|^2_{x^{2c-1}})$$

still holds true. It remains to bound the variance of our estimator. Here, we will frequently use the fact that conditioned on $\sigma(X_j, j\in[\![n]\!])$ the random variable $(Y_j)_{j\in[\![n]\!]}$ are independent but not identically distributed. Nevertheless, this allows us to split the variance into two terms, an inverse problem term, which occurs also in the i.i.d. case, compare Proposition 3.4.10 and a dependence term. This is captured in the following Proposition.

Proposition 3.6.4 (Upper bound of the risk)**:**
Let $f\in\mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$, $\sigma_c := \mathbb{E}_{f_Y}(Y_1^{2(c-1)}) < \infty$ and **[G0]** be fulfilled. Then for any $k\in\mathbb{R}_+$

$$\mathbb{E}^n_{f_Y}(\|f_k - \widehat{f}_k\|^2_{x^{2c-1}}) \leq \frac{\sigma_c}{n}\Delta_g(k) + \frac{1}{2\pi}\int_{[-k,k]}\mathbb{V}\text{ar}^n_f(\widehat{\mathcal{M}}_X(t))d\lambda(t),$$

where $\widehat{\mathcal{M}}_X(t) := n^{-1}\sum_{j=1}^n X_j^{c-1+it}$, and for $\Delta_g(k) := (2\pi)^{-1}\int_{[-k,k]}|\mathcal{M}_c[g](t)|^{-2}d\lambda(t)$ and $\widehat{f}_k$ defined in Equation (3.14). Consequently, one has

$$\mathbb{E}^n_{f_Y}(\|f - \widehat{f}_k\|^2_{x^{2c-1}}) \leq \|f - f_k\|^2_{x^{2c-1}} + \frac{\sigma_c}{n}\Delta_g(k) + \frac{1}{2\pi}\int_{[-k,k]}\mathbb{V}\text{ar}^n_f(\widehat{\mathcal{M}}_X(t))d\lambda(t).$$

Proof of Proposition 3.6.4. Let us denote by $\mathbb{E}_{|X}(\cdot) := \mathbb{E}(\cdot|X_1, \dots, X_n)$ the conditional expectation with respect to $(X_j)_{j\in[\![n]\!]}$. Then due to the independence between $(U_j)_{j\in[\![n]\!]}$ and $(X_j)_{j\in[\![n]\!]}$ we have $\mathbb{E}_{|X}(U_j^{c-1+it}) = \mathcal{M}_c[g](t)$ and thus

$$\mathbb{E}_{|X}(\widehat{\mathcal{M}}_c(t)) = \mathcal{M}_c[g](t)\widehat{\mathcal{M}}_X(t), \quad t \in \mathbb{R}.$$

By a direct calculus, we see that

$$\mathbb{E}^n_{f_Y}(|\widehat{\mathcal{M}}_c(t) - \mathcal{M}_c[f_Y](t)|^2) = \mathbb{E}^n_{f_Y}(|\widehat{\mathcal{M}}_c(t) - \mathbb{E}_{|X}(\widehat{\mathcal{M}}_c(t))|^2) + |\mathcal{M}_c[g](t)|^2 \mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t))$$

which implies, by exploiting that for $j, j' \in [\![n]\!]$

$$\mathbb{E}^n_{f_Y}(X_j^{c-1+it}\overline{X_{j'}^{c-1+it}}(U_j^{c-1+it} - \mathcal{M}_c[g](t))\overline{(U_{j'}^{c-1+it} - \mathcal{M}_c[g](t))}) = \delta_{j,j}\mathbb{E}_f(X_1^{2(c-1)})\mathbb{V}\mathrm{ar}_g(U_1^{c-1+it}),$$

here $\delta_{j,j'} = \mathbb{1}_{j=j'}$, that

$$\begin{aligned}\mathbb{E}^n_{f_Y}(|\widehat{\mathcal{M}}_c(t) - \mathcal{M}_c[f_Y](t)|^2) &= \frac{\mathbb{E}_f(X_1^{2(c-1)})\mathbb{V}\mathrm{ar}_g(U_1^{c-1+it})}{n} + |\mathcal{M}_c[g](t)|^2\mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t)) \\ &\leq \frac{\sigma_c}{n} + |\mathcal{M}_c[g](t)|^2\mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t)).\end{aligned}$$

We deduce that

$$\begin{aligned}\mathbb{E}^n_{f_Y}(\|\widehat{f}_k - f_k\|^2_{x^{2c-1}}) &= \frac{1}{2\pi}\int_{[-k,k]} \frac{\mathbb{E}^n_{f_Y}(|\widehat{\mathcal{M}}_c(t) - \mathcal{M}_c[f_Y](t)|^2)}{|\mathcal{M}_c[g](t)|^2} d\lambda(t) \\ &\leq \frac{\sigma_c\Delta_g(k)}{n} + \frac{1}{2\pi}\int_{[-k,k]} \mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t))d\lambda(t).\end{aligned}$$

The rest of the proof follows the same steps as Proposition 3.2.1. □

Remark 3.6.5 (Decomposition of the variance)**:**
In Proposition 3.6.4, we decomposed the the variance term into two terms and simplify it to

$$\mathbb{E}^n_{f_Y}(\|f_k - \widehat{f}_k\|^2_{x^{2c-1}}) \leq \frac{\sigma_c\Delta_g(k)}{n} + \frac{1}{2\pi}\int_{[-k,k]} \mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t))d\lambda(t),$$

where the first term already arises in Proposition 3.2.1 and is characterized by the decay of the Mellin transform of the error density. It can be seen as the influence of the underlying statistical inverse problem on the risk. The second summand is characterized by the dependency structure of $(X_j)_{j\in[\![n]\!]}$ and is independent of the error density g. Indeed, in the i.i.d. case, we have

$$\frac{1}{2\pi}\int_{[-k,k]} \mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t))d\lambda(t) \leq \frac{\mathbb{E}_f(X_1^{2(c-1)})}{n}\frac{2k}{2\pi} \leq \sigma_X\frac{k}{n}$$

with $\sigma_X := \mathbb{E}_f(X_1^{2(c-1)})$. Thus, in the i.i.d. case, the growth of the variance term is determined by the growth of the inverse problem term $\sigma_c\Delta_g(k)n^{-1}$.

To specify the result of Proposition 3.6.4, we will consider the case of β-mixing processes. To do so, we will be in need of the following Lemma. The key statement regarding β-mixing processes is delivered by the proposed variance bound derived by Asin and Johannes (2017) after Lemma 4.1 of the same work. Their approach is based on the original idea of Viennet (1997, Theorem 2.1).

Lemma 3.6.6 (Variance bound of β-mixing):
Let $(Z_j)_{j\in\mathbb{N}_0}$ be a strictly stationary process of real-valued random variables with common marginal distribution $\mathbb{P}$. There exists a sequence $(b_k)_{k\in\mathbb{N}}$ of measurable functions $b_k : \mathbb{R} \to [0,1]$ with $\mathbb{E}_{\mathbb{P}}[b_k(Z_0)] = \beta(Z_0, Z_k)$, such that for any measurable function h with $\mathbb{E}[|h(Z_0)|^2] < \infty$, $b = \sum_{k=1}^{\infty}(k+1)^{p-2}b_k : \mathbb{R} \to [0,\infty]$, $p \geq 2$,

$$\mathrm{Var}_{\mathbb{P}}^n\left(\sum_{j=1}^n h(Z_j)\right) \leq 4n\mathbb{E}[|h(Z_0)|^2 b(Z_0)],$$

where we set $b_0 \equiv 1$.

The proof of Lemma 3.6.6 can be found in Asin and Johannes (2017). The following Corollary is a direct consequence of Proposition 3.6.4 and Lemma 3.6.6 and is thus omitted.

Corollary 3.6.7:
Under the assumptions of Proposition 3.6.4 and for $(X_j)_{j\in[\![n]\!]}$ β-mixing with $\mathbb{E}_f(X_1^{2(c-1)}b(X_1)) < \infty$, it holds

$$\mathbb{E}_{f_Y}^n(\|f - \widehat{f}_k\|_{x^{2c-1}}^2) \leq \|f - f_k\|_{x^{2c-1}}^2 + \frac{\sigma_c}{n}\Delta_g(k) + 2\mathbb{E}_f(X_1^{2(c-1)}b(X_1))\frac{k}{n}.$$

Remark 3.6.8:
For any error distributions with $\lim_{t\to\infty}|\mathcal{M}_c[g](t)| \to 0$ we have that $k/\Delta_g(k) \to 0$, implying that in these cases, the second summand is negligable or in other words, the underlying dependence structure does not disorientate the rate since the underlying inverse problem has a more significant effect on the variance term of the estimator.

In the next part, we will discuss a fully data-driven choice of $k \in \mathbb{R}_+$. Indeed, we will present a data-driven choice for both cases, smooth and super smooth error densities.

3.6.3 Data-driven method

Let us propose a data-driven choice of $k \in \mathbb{R}_+$ in analogy to Section 3.3.3 for the super smooth case, compare Definition 3.3.1. We focus on this case, as it naturally arise while considering stochastic volatility density estimation, Section 3.6.4. To do so, we reduce the space of possible cut-off parameters to

$$\mathcal{K}_n := \{k \in [\![K_n]\!] : \Delta_g(k) \leq n\}, \quad K_n := (\log(n)/2\lambda)^{1/\rho},$$

where we assume $\rho \geq 1$. Here, the parameters $\rho, \lambda \in \mathbb{R}_+$ are taken from Definition 3.3.1. Then we define the model selection method for $\chi > 0$ and $\widehat{\sigma}_Y := \frac{1}{n}\sum_{j\in[\![n]\!]} Y_j^{2(c-1)}$

$$\widehat{k} := \arg\min_{k\in\mathcal{K}_n} -\|\widehat{f}_k\|_{x^{2c-1}}^2 + \widehat{\mathrm{pen}}(k), \quad \text{where} \breve{a}\widehat{\mathrm{pen}}(k) := \frac{\chi\widehat{\sigma}_Y k^\rho \Delta_g(k)}{n}. \tag{3.43}$$

Again, the penality term $(\mathrm{pen}(k))_{k\in\mathcal{K}_n}$ with

$$\mathrm{pen}(k) = \mathbb{E}_{f_Y}^n(\widehat{\mathrm{pen}}(k)) = \chi\sigma_Y\frac{k^\rho\Delta_g(k)}{n}, \quad k \in \mathcal{K}_n,$$

overestimates the variance term as in Section 3.3. Furthermore, the penality only reflects the influence of the inverse problem on the variance term and neglects the maybe unknown dependency structure.

Theorem 3.6.9 (Data-driven choice of $k \in \mathbb{R}_+$)**:**
Let $c \in \mathbb{R}$ with $\mathbb{E}_{f_Y}(Y_1^{4(c-1)}) < \infty$ and **[G2]** holds true with $\rho \geq 1$. Then there exists $\chi_0 \in \mathbb{R}_+$, such that for all $\chi \geq \chi_0$

$$\mathbb{E}^n_{f_Y}(\|\widehat{f}_{\widehat{k}} - f_{K_n}\|^2_{x^{2c-1}}) \leq 3 \inf_{k \in \mathcal{K}_n} \left(\|f_{K_n} - f_k\|^2_{x^{2c-1}} + \mathrm{pen}(k)\right) + \frac{C(g, \chi, \mathbb{E}_f(X_1^{4(c-1)}))}{n} + \int_{[-K_n, K_n]} \mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t)) d\lambda(t) + C(g, \chi, \sigma_X) \mathbb{V}\mathrm{ar}^n_f(\widehat{\sigma}_X) \log(n),$$

where $C(\cdot)$ is a positive constant only depending on its arguments. If further $f \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$, then

$$\mathbb{E}^n_{f_Y}(\|\widehat{f}_{\widehat{k}} - f\|^2_{x^{2c-1}}) \leq 3 \inf_{k \in \mathcal{K}_n} \left(\|f - f_k\|^2_{x^{2c-1}} + \mathrm{pen}(k)\right) + \frac{C(g, \chi, \mathbb{E}_f(X_1^{4(c-1)}))}{n} + \int_{[-K_n, K_n]} \mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t)) d\lambda(t) + C(g, \chi, \sigma_X) \mathbb{V}\mathrm{ar}^n_f(\widehat{\sigma}_X) \log(n).$$

Remark 3.6.10 (Comment on the additional dependency term)**:**
Comparing to the result of Proposition 3.6.4, an additional dependency term, $\mathbb{V}\mathrm{ar}^n_f(\widehat{\sigma}_X)\log(n)$, arises in the bounds of Theorem 3.6.9. Indeed, this dependency term arises as a direct consequence of the need of the estimation of σ_c. For $c = 1$ the estimation of the moment is not necessary, which is why this additional term is not needed. Furthermore, the log-term in the super smooth case is due to the overestimation of the variance term. In the smooth error case the log-term could be omitted.

Proof of Theorem 3.6.9. First, since $\mathbb{E}_{f_Y}(Y_1^{c-1}) < \infty$ we have that $\mathcal{M}_c[f]$, $\mathcal{M}_c[g]$ and $\mathcal{M}_c[f_Y]$ are well-defined, and so $(f_k)_{k \in \mathcal{K}_n}$ even without the assumption $f \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$. Now we can state the following Lemma whose proof can be found in the proof section.

Lemma 3.6.11:
Under the assumptions of Theorem 3.6.9, it holds

$$\mathbb{E}^n_{f_Y}(\|\widehat{f}_{\widehat{k}} - f_{K_n}\|^2_{x^{2c-1}}) \leq 3\left(\|f_{K_n} - f_k\|^2_{x^{2c-1}} + \mathrm{pen}(k)\right) + 11\mathbb{E}^n_{f_Y}\left(\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}} - \frac{1}{12}\mathrm{pen}(\widehat{k})\right)_+ + \mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - 2\widehat{\mathrm{pen}}(\widehat{k}))_+) + \int_{[-K_n, K_n]} \mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t)) d\lambda(t).$$

The first inequality of the theorem follows by applying the following tow Lemmas and taking the infimum over $k \in \mathcal{K}_n$.

Lemma 3.6.12:
Under the assumptions of Theorem 3.6.9 we get

$$\mathbb{E}^n_{f_Y}\left(\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}} - \frac{1}{12}\mathrm{pen}(\widehat{k})\right)_+ \leq \frac{C(g, \mathbb{E}_f(X_1^{4(c-1)}))}{n} + \int_{[-K_n,K_n]} \mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t))d\lambda(t)$$
$$+ C(g, \chi, \sigma_X)\mathbb{V}\mathrm{ar}^n_f(\widehat{\sigma}_X)\log(n),$$

where $C(\cdot) > 0$ is a positive constant only depending only on its arguments.

Lemma 3.6.13:
Under the assumptions of Theorem 3.6.9 we get

$$\mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - 2\widehat{\mathrm{pen}}(\widehat{k}))_+) \leq C(\chi, \sigma_Y, \mathbb{E}_{f_Y}(Y_1^{4(c-1)}))\log(n)(n^{-1} + \mathbb{V}\mathrm{ar}^n_f(\widehat{\sigma}_X)),$$

where $C(\cdot) > 0$ is a positive constant only depending only on its arguments.

The second inequality is a direct consequence of $\|f - \widehat{f}_k\|^2_{x^{2c-1}} = \|f - f_{K_n}\|^2_{x^{2c-1}} + \|f_{K_n} - \widehat{f}_k\|^2_{x^{2c-1}}$ for any $k \in \mathcal{K}_n$ including $\widehat{k}$. □

We will now give the proofs of Lemma 3.6.12 and Lemma 3.6.13 to demonstrate the usage of the conditional expectation in combination with the Talagrand inequality.

Proof of Lemma 3.6.12. First we see that

$$\mathbb{E}^n_{f_Y}\left(\left(\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}} - \frac{1}{12}\mathrm{pen}(\widehat{k})\right)_+\right) \leq \mathbb{E}^n_{f_Y}\left(\max_{k\in\mathcal{K}_n}\left(\|\widehat{f}_k - f_k\|^2_{x^{2c-1}} - \frac{1}{12}\mathrm{pen}(k)\right)_+\right).$$

Define $B_k := \{h \in S_k : \|h\|_{x^{2c-1}} = 1\}$. Furthermore, we define

$$\|\widehat{f}_k - f_k\|^2_{x^{2c-1}} = \sup_{h\in B_k} \langle \widehat{f}_k - f_k, h\rangle^2_{x^{2c-1}} =: \sup_{h\in B_k} \overline{\nu}_h^2$$

and decompose $\overline{\nu}_h$ into $\overline{\nu}_h = \overline{\nu}_{h,in} + \overline{\nu}_{h,de}$, where

$$\overline{\nu}_{h,in} := n^{-1}\sum_{j\in[\![n]\!]}(\nu_h(Y_j) - \mathbb{E}_{|X}(\nu_h(Y_j))), \quad \overline{\nu}_{h,de} = n^{-1}\sum_{j\in[\![n]\!]}\mathbb{E}_{|X}(\nu_h(Y_j))) - \mathbb{E}^n_{f_Y}(\nu_h(Y_j))$$

and

$$\nu_h(Y_j) := \frac{1}{2\pi}\int_{[-k,k]}\frac{Y_j^{c-1+it}}{\mathcal{M}_c[g](t)}\mathcal{M}_c[h](t)d\lambda(t).$$

Thus

$$\mathbb{E}^n_{f_Y}\left(\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}} - \frac{1}{12}\mathrm{pen}(\widehat{k})\right)_+ \leq 2\mathbb{E}^n_{f_Y}\left(\max_{k\in\mathcal{K}_n}\left(\overline{\nu}^2_{h,in} - \frac{1}{24}\mathrm{pen}(k)\right)_+\right) + 2\mathbb{E}^n_{f_Y}\left(\max_{k\in\mathcal{K}_n}\overline{\nu}^2_{h,de}\right).$$

Starting with the second dependent term, applying the Cauchy-Schwarz inequality and the fact $\mathbb{E}_{|X}(Y_j^{c-1+it}) - \mathbb{E}_{f_Y}(Y_j^{c-1+it}) = \mathcal{M}_c[g](t)(X_j^{c-1+it} - \mathbb{E}_f(X_1^{c-1+it})$ we get for any $k \in \mathcal{K}_n$ and any

$h \in B_k$,

$$\left|\sum_{j\in\llbracket n\rrbracket} \mathbb{E}_{|X}(\nu_h(Y_j)) - \mathbb{E}_{f_Y}(\nu_h(Y_j))\right|^2 = \frac{1}{4\pi^2}\left|\int_{[-k,k]}\left(\sum_{j\in\llbracket n\rrbracket} X_j^{c-1+it} - \mathbb{E}_f(X_j^{c-1+it})\right)\overline{\mathcal{M}_c[h](t)}d\lambda(t)\right|^2$$
$$\leq \frac{\|h\|^2_{x^{2c-1}}}{2\pi}\int_{[-k,k]}\left|\sum_{j\in\llbracket n\rrbracket} X_j^{c-1+it} - \mathbb{E}_f(X_j^{c-1+it})\right|^2 d\lambda(t).$$

Since $\|h\|_{x^{2c-1}} \leq 1$, we get for the dependent term

$$\mathbb{E}^n_{f_Y}(\max_{k\in\mathcal{K}_n}\sup_{h\in B_k}\bar{\nu}^2_{h,de}) \leq \frac{1}{2\pi}\int_{[-K_n,K_n]}\mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t))d\lambda(t).$$

Next, let us define $\widetilde{\mathrm{pen}}(k) := \mathbb{E}_{|X}(\widehat{\mathrm{pen}}(k)) = \chi\sigma_U\widehat{\sigma}_X k^\rho\Delta_g(k)n^{-1}$. Then,

$$\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}\bar{\nu}^2_{h,in} - \frac{1}{24}\mathrm{pen}(k))_+ = \max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}\bar{\nu}^2_{h,in} - \frac{1}{36}\widetilde{\mathrm{pen}}(k))_+ + \frac{1}{24}\max_{k\in\mathcal{K}_n}(\frac{2}{3}\widetilde{\mathrm{pen}}(k) - \mathrm{pen}(k))_+.$$

For the second summand, we have

$$\mathbb{E}^n_{f_Y}(\max_{k\in\mathcal{K}_n}(\widetilde{\mathrm{pen}}(k) - \mathrm{pen}(k))_+) \leq \chi K_n^\rho\sigma_U\mathbb{E}^n_f((\frac{2}{3}\widehat{\sigma}_X - \sigma_X)_+).$$

Let us define $\Omega_X := \{|\widehat{\sigma}_X - \sigma_X| \leq \sigma_X/2\}$. Then on Ω_X we have $\widehat{\sigma}_X \leq 3\sigma_X/2$ and thus $\mathbb{E}^n_f((\frac{2}{3}\widehat{\sigma}_X - \sigma_X)_+) \leq \mathbb{E}^n_f((\widehat{\sigma}_X - \sigma_X)_+\mathbb{1}_{\Omega^c_X}) \leq 2\sigma_X^{-1}\mathbb{V}\mathrm{ar}^n_f(\widehat{\sigma}_X)$ by application of the Cauchy-Schwarz and the Markov's inequality. This implies

$$\mathbb{E}^n_{f_Y}(\max_{k\in\mathcal{K}_n}(\widetilde{\mathrm{pen}}(k) - \mathrm{pen}(k))_+) \leq C(g,\chi,\sigma_X)\log(n)\mathbb{V}\mathrm{ar}^n_f(\widehat{\sigma}_X).$$

For the first summand, we see

$$\mathbb{E}^n_{f_Y}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}\bar{\nu}^2_{h,in} - \frac{1}{36}\widetilde{\mathrm{pen}}(k))_+) = \mathbb{E}^n_f(\mathbb{E}_{|X}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}\bar{\nu}^2_{h,in} - \frac{1}{36}\widetilde{\mathrm{pen}}(k))_+)).$$

Thus, we start by considering the inner conditional expectation to bound the term. By the construction of $\bar{\nu}_{h,in}$, its summands conditioned on $\sigma(X_i, i \geq 0)$ are independent but not identically distributed. We are aiming to apply the Talagrand inequality, Lemma 3.3.12. We therefore split, for a sequence $(d_n)_{n\in\mathbb{N}}$ specified afterwards, the process again in the following way

$$\bar{\nu}_{h,1} := n^{-1}\sum_{j\in\llbracket n\rrbracket}\nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{c-1}) - \mathbb{E}_{|X}(\nu_h(Y_1)\mathbb{1}_{(0,d_n)}(Y_j^{c-1}))$$
$$\text{and } \bar{\nu}_{h,2} := n^{-1}\sum_{j\in\llbracket n\rrbracket}\nu_h(Y_j)\mathbb{1}_{(d_n,\infty)}(Y_j^{c-1}) - \mathbb{E}_{|X}(\nu_h(Y_j)\mathbb{1}_{(d_n,\infty)}(Y_j^{c-1}))$$

to get

$$\mathbb{E}_{|X}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}|\bar{\nu}_{h,in}|^2 - \frac{1}{36}\widetilde{\mathrm{pen}}(k))_+) \leq 2\mathbb{E}_{|X}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}|\bar{\nu}_{h,1}|^2 - \frac{1}{72}\widetilde{\mathrm{pen}}(k))_+ + |\bar{\nu}_{h,2}|^2),$$
$$:= M_1 + M_2.$$

We will now consider the two summands M_1, M_2 separately.
To bound M_1 we will use the Talagrand inequality, Lemma 3.3.12, on $\mathbb{E}_{|X}(\sup_{t\in B_k}|\bar{\nu}_{h,1}|^2 - \frac{1}{72}\widetilde{\mathrm{pen}}(k))_+$. Indeed, we have

$$M_1 \leq \sum_{k\leq K_n} \mathbb{E}_{|X}(\sup_{t\in B_k}|\bar{\nu}_{h,1}|^2 - \frac{1}{72}\widetilde{\mathrm{pen}}(k))_+,$$

which will be used to show the claim. We want to emphasize that we are able to apply the Talagrand inequality on the sets B_k, since B_k has a dense countable subset and due to continuity arguments. Further, we see that the random variables $\nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{c-1}) - \mathbb{E}_{|X}(\nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{c-1}))$, $j\in [\![n]\!]$, are, conditioned on $\sigma(X_i, i\geq 0)$, centered and independent but not identically distributed. In order to apply Talagrand's inequality, we need to find the constants Ψ, ψ, τ such that

$$\sup_{h\in B_k}\sup_{y\in\mathbb{R}_+}|\nu_h(y)\mathbb{1}_{(0,d_n)}(y^{c-1})| \leq \psi; \quad \mathbb{E}_{|X}(\sup_{h\in B_k}|\bar{\nu}_{h,1}|) \leq \Psi;$$
$$\sup_{h\in B_k}\frac{1}{n}\sum_{j\in[\![n]\!]}\mathbb{V}\mathrm{ar}_{|X}(\nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{c-1})) \leq \tau.$$

We start to determine the constant Ψ^2. Let us define $\widetilde{\mathcal{M}}_c(t) := n^{-1}\sum_{j\in[\![n]\!]} Y_j^{c-1+it}\mathbb{1}_{(0,d_n)}(Y_j^{c-1})$ as an unbiased estimator of $\mathcal{M}_c[f_Y\mathbb{1}_{(0,d_n)}(y^{c-1})](t)$ and

$$\widetilde{f}_k(x) := \frac{1}{2\pi}\int_{[-k,k]} x^{-c-it}\frac{\widetilde{\mathcal{M}}(t)}{\mathcal{M}_c[g](t)}d\lambda(t),$$

where $n^{-1}\sum_{j\in[\![n]\!]}\nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{c-1}) = \langle \widetilde{f}_k, h\rangle_{x^{2c-1}}$. Thus, we have for any $h\in B_k$ that $\bar{\nu}_{h,1}^2 = \langle h, \widetilde{f}_k - \mathbb{E}_{|X}(\widetilde{f}_k)\rangle^2_{x^{2c-1}} \leq \|h\|^2_{x^{2c-1}}\|\widetilde{f}_k - \mathbb{E}_{|X}(\widetilde{f}_k)\|^2_{x^{2c-1}}$. Since $\|h\|_{x^{2c-1}}\leq 1$, we get

$$\mathbb{E}_{|X}(\sup_{h\in B_k}\bar{\nu}_{h,1}^2) \leq \mathbb{E}_{|X}(\|\widetilde{f}_k - \mathbb{E}_{|X}(\widetilde{f}_k)\|^2) = \frac{1}{2\pi}\int_{[-k,k]}\frac{\mathbb{E}_{|X}(|\widetilde{\mathcal{M}}(t) - \mathbb{E}_{|X}(\widetilde{\mathcal{M}}(t))|^2)}{|\mathcal{M}_c[g](t)|^2}d\lambda(t).$$

Now, since $Y_j^{c-1+it}\mathbb{1}_{(0,d_n)}(Y_j^{c-1}) - \mathbb{E}_{|X}(Y_j^{c-1+it}\mathbb{1}_{(0,d_n)}(Y_j^{c-1})$ are independent, conditioned on $\sigma(X_i : i\geq 0)$, we obtain

$$\mathbb{E}_{|X}(|\widetilde{\mathcal{M}}(t) - \mathbb{E}_{|X}(\widetilde{\mathcal{M}}(t))|^2) \leq \frac{1}{n^2}\sum_{j\in[\![n]\!]}\mathbb{E}_{|X}(Y_j^{2(c-1)}\mathbb{1}_{(0,d_n)}(Y_j^{c-1})) = \frac{\sigma_U}{n}\widehat{\sigma}_X,$$

which implies

$$\mathbb{E}_{|X}(\sup_{h\in B_k}\bar{\nu}_{h,1}^2) \leq \sigma_U\widehat{\sigma}_X\frac{\Delta_g(k)}{n} =: \Psi^2.$$

Next we consider ψ. Let $y\in\mathbb{R}_+$ and $h\in B_k$. Then using the Cauchy-Schwarz inequality,

$$|\nu_h(y)\mathbb{1}_{(0,d_n)}(y^{c-1})|^2 = (2\pi)^{-2}d_n^2|\int_{[-k,k]} y^{it}\frac{\mathcal{M}_c[h](-t)}{\mathcal{M}_c[g](t)}d\lambda(t)|^2$$
$$\leq (2\pi)^{-2}d_n^2\int_{[-k,k]}|\mathcal{M}_c[g](t)|^{-2}d\lambda(t) \leq d_n^2\Delta_g(k) =: \psi^2$$

, since $|y^{it}| = 1$ for all $t \in \mathbb{R}$. For τ we use the crude bound $\tau = n\Psi^2$. Hence, we have $\frac{n\Psi^2}{\tau} = 1$, $\frac{n\Psi}{\psi} = \frac{\sqrt{\sigma_U \widehat{\sigma}_X n}}{d_n}$ and get

$$\mathbb{E}_{|X}\big(\sup_{h\in B_k} \bar{\nu}_{h,1}^2 - 2(1+2\varepsilon)\sigma_U\widehat{\sigma}_X\frac{\Delta_g(k)}{n}\big)_+ \leq \frac{C}{n}\Big(\widehat{\sigma}_X\sigma_U\Delta_g(k)\exp(-K_1\varepsilon) + \frac{\Delta_g(k)d_n^2}{n}\exp(-K_2C_\varepsilon\sqrt{\varepsilon}\sqrt{\sigma_U\widehat{\sigma}_X n}d_n^{-1})\Big).$$

Choosing now $\varepsilon = 4\lambda k^\rho/K_1$ and t applying assumption **[G2]** for $k \geq k_g$, we get

$$\Delta_g(k)\exp(-K_1\varepsilon) \leq C_g k^{2\gamma}\exp(-\lambda k^\rho),$$

which is summable over $\mathbb{N}$. Next for $k \geq k_g$ we get $C_\varepsilon\sqrt{\varepsilon} \geq \varepsilon/2$ and choosing $d_n := \sqrt{n\sigma_U\widehat{\sigma}_X}2K_2/K_1$ leads to

$$\begin{aligned}\mathbb{E}_{|X}\big(\sup_{h\in B_k} \bar{\nu}_{h,1}^2 - 2(1+2\varepsilon)\sigma_U\widehat{\sigma}_X\frac{\Delta_g(k)}{n}\big)_+ &\leq \frac{C_g}{n}\widehat{\sigma}_X\sigma_U\left(k^{2\gamma}\exp(-\lambda k^\rho) + \Delta_g(k)\exp(-K_1\varepsilon)\right)\\ &\leq \frac{C_g}{n}\widehat{\sigma}_X\sigma_U k^{2\gamma}\exp(-\lambda k^\rho).\end{aligned}$$

Hence, there exists a $\chi_0 > 0$, such that for all $\chi > \chi_0$ holds $\frac{1}{72}\widetilde{\text{pen}}(k) \geq 2(1+4\lambda k^\rho)/K_1)\sigma_U\widehat{\sigma}_X\Delta_g(k)n^{-1}$ implying

$$\sum_{k\leq K_n}\mathbb{E}_{|X}(\sup_{t\in B_k}|\bar{\nu}_{h,1}|^2 - \frac{1}{72}\widetilde{\text{pen}}(k))_+ \leq \frac{C_g}{n}\widehat{\sigma}_X\sigma_U\sum_{k\leq K_n}k^{2\gamma}\exp(-\lambda k^\rho) \leq \frac{C(g)\widehat{\sigma}_X}{n}.$$

Now, we consider M_2. Let us define $\overline{f}_k := \widehat{f}_k - \widetilde{f}_k$. Then from $\overline{\nu}_{h,2} = \overline{\nu}_{h,in} - \overline{\nu}_{h,1}$ we deduce $\overline{\nu}_{h,2}^2 = \langle\overline{f}_k - \mathbb{E}_{|X}(\overline{f}_k), h\rangle_{x^{2c-1}}^2 \leq \|\overline{f}_k - \mathbb{E}_{|X}(\overline{f}_k)\|_{x^{2c-1}}^2$ for any $h \in B_k$. Further,

$$\max_{k\in\mathcal{K}_n}\|\overline{f}_k - \mathbb{E}_{|X}(\overline{f}_k)\|_{x^{2c-1}}^2 \leq \|\overline{f}_{K_n} - \mathbb{E}_{|X}(\overline{f}_{K_n})\|_{x^{2c-1}}^2$$

and

$$\begin{aligned}\mathbb{E}_{|X}(\|\overline{f}_{K_n} - \mathbb{E}_{|X}(\overline{f}_{K_n}))\|_{x^{2c-1}}^2) &= \frac{1}{2\pi}\int_{[-K_n,K_n]}\frac{\mathbb{V}\text{ar}_{|X}(\widehat{\mathcal{M}}(t) - \widetilde{\mathcal{M}}(t))}{\mathcal{M}_c[g](t)|^2}d\lambda(t)\\ &\leq \frac{1}{n^2}\sum_{j=1}^n\mathbb{E}_{|X}(Y_j^{2(c-1)}\mathbb{1}_{(d_n,\infty)}(Y_j^{c-1}))\Delta_g(K_n).\end{aligned}$$

Let us define the event $\Xi_X := \{\widehat{\sigma}_X \geq \sigma_X/2\}$. Then, we have

$$\frac{1}{n^2}\sum_{j\in[\![n]\!]}\mathbb{E}_{|X}(Y_j^{2(c-1)}\mathbb{1}_{(d_n,\infty)}(Y_j^{(c-1)}))\Delta_g(K_n)\mathbb{1}_{\Xi_X} \leq \frac{C_g}{n}\sum_{j\in[\![n]\!]}X_j^{(2+p)(c-1)}\mathbb{E}_g(U_j^{(2+p)(c-1)})d_n^{-p}\mathbb{1}_{\Xi_X},$$

where on Ξ_X we can state that $d_n^{-p} = C(g)n^{-p/2}(\widehat{\sigma}_X)^{-p/2} \leq C(g)\sigma_X^{-p/2}n^{-p/2}$. Choosing $p = 2$ leads to $\mathbb{E}_{|X}(\|\overline{f}_{K_n} - \mathbb{E}_{|X}(\overline{f}_{K_n}))\|_{x^{2c-1}}^2)\mathbb{1}_{\Xi_X} \leq \frac{C(g)\sigma_X^{-1}}{n}\mathbb{E}_g(U_1^{4(c-1)})n^{-1}\sum_{j=1}^n X_j^{4(c-1)}$. On the other hand,

$$\frac{1}{n^2}\sum_{j=1}^n\mathbb{E}_{|X}(Y_j^{2(c-1)}\mathbb{1}_{(d_n,\infty)}(Y_j^{c-1}))\Delta_g(k)\mathbb{1}_{\Xi_X^c} \leq \frac{\sigma_Y}{2}\mathbb{1}_{\Xi_X^c} \leq \frac{\sigma_Y}{2}\mathbb{1}_{\Omega_X^c}.$$

We get

$$M_2 \leq C(g, \sigma_X) \left(\frac{\mathbb{E}_f(X_1^{4(c-1)})}{n} + \mathbb{V}\mathrm{ar}_f^n(\widehat{\sigma}_X) \right).$$

These three bounds imply

$$\mathbb{E}_{f_Y}^n \left(\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|_{x^{2c-1}}^2 - \frac{1}{12}\mathrm{pen}(\widehat{k}) \right)_+ \leq \frac{C(g, \mathbb{E}_f(X_1^{4(c-1)}))}{n} + C(g, \chi, \sigma_X)\mathbb{V}\mathrm{ar}_f^n(\widehat{\sigma}_X)\log(n).$$

□

Proof of Lemma 3.6.13. Let us define $\Omega := \{|\widehat{\sigma}_Y - \sigma_Y| \leq \sigma_Y/2\}$. Then on Ω we have $2\widehat{\sigma}_Y \geq \sigma_Y$, s.t.

$$\begin{aligned} \mathbb{E}_{f_Y}^n((\mathrm{pen}(k) - 2\widehat{\mathrm{pen}}(k))_+) &= \chi \mathbb{E}_{f_Y}^n \left(\widehat{k}\frac{\Delta_g(\widehat{k}^\rho)}{n}(\sigma_Y - 2\widehat{\sigma}_Y)_+ \right) \\ &\leq 2\chi K_n^\rho \mathbb{E}_{f_Y}^n(|\sigma_Y - \widehat{\sigma}_Y|\mathbb{1}_{\Omega^c}) \\ &\leq 2\chi \log(n)\frac{\mathbb{V}\mathrm{ar}_{f_Y}^n(\widehat{\sigma}_Y)}{\sigma_Y}. \end{aligned}$$

Now in analogy to the proof of Proposition 3.6.4, we get

$$\mathbb{V}\mathrm{ar}_{f_Y}^n(\widehat{\sigma}_Y) = \frac{\mathbb{E}_{f_Y}(Y_1^{4(c-1)})}{n} + \mathbb{E}_g(U_1^{2(c-1)})^2\mathbb{V}\mathrm{ar}_f^n(\widehat{\sigma}_X).$$

□

As we have seen, the Theorem 3.6.9 implies, that the fully data-driven estimator $\widehat{f}_{\widehat{k}}$ proposed in Section 3.3.3 has a reasonable upper bound even in the dependent data case. It is worth stressing out, that the estimator is not dependent on the knowledge if the data possesses a dependence structure or not. We will now see, that in the case of β-mixing dependency, we can still achieve an oracle inequality with a negliable additive term.

Corollary 3.6.14 (Data-driven choice for β-mixing)**:**
Under the assumptions of Theorem 3.6.9 and for $(X_j)_{j\in[\![n]\!]}$ β-mixing with $\mathbb{E}_f(X_1^{4(c-1)}b(X_1)) < \infty$ holds

$$\mathbb{E}_{f_Y}^n(\|\widehat{f}_{\widehat{k}} - f\|_{x^{2c-1}}^2) \leq 3 \inf_{k\in\mathcal{K}_n} \left(\|f - f_k\|_{x^{2c-1}}^2 + \mathrm{pen}(k)\right) + C(g, \chi, f)\frac{\log(n)}{n},$$

where $C(g, \chi, f)$ is a positive constant depending on $\mathbb{E}_{f_Y}(Y_1^{4(c-1)}), \sigma_Y, \mathbb{E}_f(X_1^{4(c-1)}b(X_1))$ and χ.

As we have seen in Section 3.3.3 the choice of $\mathcal{K}_n$ is not restrictive. In the next section, we introduce the stochastic volatility model and we will see, that the developed theory for the multiplicative deconvolution for stationary process can be applied in this scenario.

3.6.4 Digression: Stochastic volatility models

Let us consider the stochastic volatility model given by

$$dZ_t = \sqrt{V_t} dW_t, \quad Z_0 = 0, \tag{3.44}$$

where $(V_t)_{t\geq 0}$ is the unique solution of the stochastic differential equation

$$dV_t = b(V_t)dt + a(V_t)dB_t, \quad V_0 = \eta, \tag{3.45}$$

and $(W_t, B_t)_{t\geq 0}$ is a standard Brownian motion, η a positive random variable independent of $(W_t, B_t)_{t\geq 0}$ and b, a are deterministic functions, such that the stochastic differential equation (3.45) possesses a unique and strictly stationary solution. The unobserved positive Markov process $(V_t)_{t\geq 0}$ is the so-called volatility process.
We are interested the case, where we observe the process $(Z_t)_{t\geq 0}$ in a discrete time and high frequency regime. Here, discrete times means that we observe the process $(Z_t)_{t\geq 0}$ only for a discrete subset $\mathbb{T} = \{t_0 \leq t_1 \leq \cdots \leq t_q\} \subset \mathbb{R}_+$. The high frequency regimes expresses that we are interested in the behaviour of our estimator for $\sup_{j\in[\![q]\!]} |t_j - t_{j-1}| \to 0$. Here, we assume that we have access to a sample $(Z_{j\Delta})_{j\in[\![n]\!]}$ drawn from (3.44).
In this section, we are interested in estimating the density $f_V : \mathbb{R}_+ \to \mathbb{R}_+$ of a strictly stationary process $(V_t)_{t\geq 0}$ given the sample $(Z_{j\Delta})_{j\in[\![n]\!]}$ from the stochastic volatility model.
This model has been consider by Genon-Catalot et al. (2003), respectively Comte and Genon-Catalot (2006). Based on their approach we can identify an underlying multiplicative deconvolution problem in the following way. For the rescaled increments

$$\frac{Z_{j\Delta} - Z_{(j-1)\Delta}}{\sqrt{\Delta}} = \frac{1}{\sqrt{\Delta}} \int_{(j-1)\Delta}^{j\Delta} V_s dW_s, \quad j \in [\![n]\!],$$

we see that conditioned on $\sigma(V_t : t \geq 0)$, the σ-field generated by $(V_t)_{t\geq 0}$, the increments are independently standard normal distributed leading to

$$\frac{Z_{j\Delta} - Z_{(j-1)\Delta}}{\sqrt{\Delta}} \sim \mathrm{N}_{(0,\overline{V}_j)} \quad \text{withă}\overline{V}_j := \frac{1}{\Delta} \int_{(j-1)\Delta}^{j\Delta} V_s ds.$$

From this, we see that the multiplicative deconvolution problem is given through

$$Y_j := \left(\frac{Z_{j\Delta} - Z_{(j-1)\Delta}}{\sqrt{\Delta}} \right)^2 = \overline{V}_j \cdot U_j, \tag{3.46}$$

where $(U_j)_{j\in[\![n]\!]}$ are i.i.d. $\Gamma_{(1/2,1/2)}$-distributed random variables.

Remark 3.6.15 (Continuous time observation)**:**
Alternatively to the discrete time observations, one could consider continuous time observations, that is we have access to $(Z_t)_{t\in[0,T]}$ for some $T > 0$. Under regularity assumption on the volatility process $(V_t)_{t\geq 0}$, the quadratic variation process $(\langle Z\rangle_t)_{t\in[0,T]}$ is observable and given through

$$\langle Z\rangle_t = \int_0^t V_s ds, \quad \text{for } t \in [0, T],$$

from which we can directly derive the paths $(V_t)_{t\in[0,T]}$. Thus, the estimation problem can be expressed as a density estimation without multiplicative measurement errors.

It is worth stressing out, that for a strictly stationary processes $(V_t)_{t\geq 0}$ the random variable $(\overline{V}_j)_{j\in[\![n]\!]}$ are again strictly stationary leading to a similar model as the general model considered in Section 3.6. Nevertheless, the stochastic volatility model differs in the way that $\overline{V}_1$ does not possesses the Lebesgue density f_V, which will lead to an additional bias term as we will see through out this digression. Before we start to bound the risk of the estimator, we collect the regularity properties, which will be assumed for the volatility process $(V_t)_{t\geq 0}$ and consider examples, which fulfills them. Throughout this section, we assume that

[A0] $(W_t, B_t)_{t\geq 0}$ in (3.44), respectively (3.45), is a two-dimensional Brownian motion stochastically independent of η and

[A1] (3.45) possesses a unique solution $(V_t)_{t\geq 0}$ which is a time-homogeneous Markov process on $\mathbb{R}_+$ with continuous sample paths, strictly stationary and ergodic. The stationary distribution of $V_0 = \eta$ admits a density f_V with respect to the Lebesgue measure on $\mathbb{R}_+$.

The assumption **[A0]** and **[A1]** imply that the sample $(\overline{V}_j)_{j\in[\![n]\!]}$, from (3.46), is a stationary process independent of $(U_j)_{j\in[\![n]\!]}$, where $U_1 \sim \Gamma_{(1/2,1/2)}$. Here we want to stress out, that $\overline{V}_1$ does not necessarily possesses a Lebesgue density, which is not needed for the upcoming theory. Let us now consider the estimator for $k \in \mathbb{R}_+$

$$\widehat{f}_k(x) := \frac{1}{2\pi}\int_{[-k,k]} x^{-1-it}\frac{\widehat{\mathcal{M}}(t)}{\mathcal{M}_1[g](t)}d\lambda(t), \quad x \in \mathbb{R}_+, \tag{3.47}$$

where $\widehat{\mathcal{M}}(t) := \widehat{\mathcal{M}}_1(t) := n^{-1}\sum_{j\in[\![n]\!]} Y_j^{it}$ and $\mathcal{M}_1[g]$ is the Mellin transform of U_1. Here, we considered the estimator (3.14) for $c = 1$ and remark later on for which $c \in \mathbb{R}$ we can expand the presented results. Due to the strict stationarity of $(\overline{V}_j)_{j\in[\![n]\!]}$, we get that for any $k \in \mathbb{R}_+$ the estimator $\widehat{f}_k$ is an unbiased estimator for

$$f_{\Delta,k}(x) := \mathbb{E}^n_{f_Y}(\widehat{f}_k(x)) = \frac{1}{2\pi}\int_{[-k,k]} x^{-1-it}\mathbb{E}(\overline{V}_1^{it})d\lambda(t), \quad x \in \mathbb{R}_+.$$

We can now state the first intermediate result concerning the variance term of our proposed estimation strategy. It is a direct consequence of Proposition 3.6.4, thus its proof is omitted.

Lemma 3.6.16 (Decomposition of the Variance term)**:**
Let **[A0]** and **[A1]** hold true. Then for $k, \Delta \in \mathbb{R}_+$

$$\mathbb{E}^n_{f_Y}(\|\widehat{f}_k - f_{\Delta,k}\|_x^2) \leq \frac{\Lambda_g(k)}{n} + \frac{1}{2\pi}\int_{[-k,k]} \mathbb{V}\mathrm{ar}(\widehat{\mathcal{M}}_{\overline{V}}(t))d\lambda(t),$$

where $\widehat{\mathcal{M}}_{\overline{V}}(t) := n^{-1}\sum_{j\in[\![n]\!]} \overline{V}_j^{it}, t \in \mathbb{R}$ and $\Lambda_g(k) := (2\pi)^{-1}\int_{[-k,k]} |\mathcal{M}_1[g](t)|^{-2}d\lambda(t)$.

Since $U_1 \sim \Gamma_{(1/2,1/2)}$, we get for $t \in \mathbb{R}$ that $\mathcal{M}_1[g](t) = 2^{it}\Gamma(1/2+it)\pi^{-1/2}$, compare Example 2.1.7, and thus

$$\Lambda_g(k) = \frac{1}{2}\int_{[-k,k]} |\Gamma(1/2+it)|^{-2}d\lambda(t) = \frac{1}{2\pi}\int_{[-k,k]} \cosh(t\pi)\lambda(t) = \frac{1}{2\pi^2}(\exp(\pi k) - \exp(-\pi k)),$$

where $\cosh(x) = (\exp(x) + \exp(-x))/2, x \in \mathbb{R}$, is the usual cosines hyperbolicus. At the same time, the family of functions $(f_{\Delta,k})_{(k,\Delta)\in\mathbb{R}_+^2}$ are approximations of f_V with $f_{\Delta,k} \in \mathbb{L}^2(\mathbb{R}_+, x)$

by construction. The approximation error, respectively the squared bias, can be bound in the following way. As before, we set for $f_V \in \mathbb{L}^2(\mathbb{R}_+, x)$ and $k \in \mathbb{R}_+$ $f_{V,k} := \mathcal{M}_1^\dagger[\mathcal{M}_1[f_Y]\mathbb{1}_{[-k,k]}]$.

Lemma 3.6.17 (Decomposition of the Bias)**:**
Let $f_V \in \mathbb{L}^2(\mathbb{R}_+, x)$ and $k, \Delta \in \mathbb{R}_+$. Then

$$\begin{aligned}\|f_V - f_{\Delta,k}\|_x^2 &= \|f_V - f_{V,k}\|_x^2 + \|f_{V,k} - f_{\Delta,k}\|_x^2 \\ &= \|f_V - f_{V,k}\|_x^2 + \frac{1}{2\pi}\int_{[-k,k]} |\mathbb{E}(V_0^{it} - \overline{V}_1^{it})|^2 d\lambda(t).\end{aligned}$$

The proof of Lemma 3.6.17 consists only of using the Pythagoras formula in the first step and Parseval identity in the second and is thus omitted.
We will now give two additional assumption on the volatility process motivated by Lemma 3.6.16 and Lemma 3.6.17.

[A2] Let $(V_t)_{t\geq 0}$ be β-mixing with $\int_{\mathbb{R}_+} \beta(t)dt < \infty$, where $\beta(t) = \beta^{\text{mix}}(\sigma(V_0), \sigma(V_t))$.

[A3] Let there exists $\mathfrak{c} > 0$, such that $\mathbb{E}(|\log(\overline{V}_1) - \log(V_0)|) \leq \mathfrak{c}\sqrt{\Delta}$.

Under assumption **[A0]**-**[A3]** we state the following result.

Theorem 3.6.18 (Upper bound of the risk)**:**
Let $f_V \in \mathbb{L}^2(\mathbb{R}_+, x)$ and **[A0]**-**[A3]** hold true. Then for any $k \in \mathbb{R}_+$ and $\Delta \in (0,1)$ holds

$$\mathbb{E}_{f_V}^n(\|f_V - \widehat{f}_k\|_x^2) \leq \|f_V - f_{V,k}\|_x^2 + \mathfrak{c}^2\Delta k^3 + \frac{\exp(\pi k)}{n} + \frac{2k}{n\Delta}\left(1 + \int_{\mathbb{R}_+} \beta(t)dt\right).$$

Proof of Theorem 3.6.18. Since $\|f_V - \widehat{f}_k\|_x^2 = \|f_V - f_{V,k}\|_x^2 + \|f_{V,k} - \widehat{f}_k\|_x^2$ it remains to show that

$$\frac{1}{2\pi}\int_{[-k,k]} \mathbb{V}\text{ar}(\widehat{\mathcal{M}}_{\overline{V}}(t))d\lambda(t) \leq \frac{k}{n\Delta}\int_{\mathbb{R}_+} \beta(t)dt \text{ and} \qquad (*)$$

$$\frac{1}{2\pi}\int_{[-k,k]} |\mathbb{E}(V_0^{it} - \overline{V}_1^{it})|^2 d\lambda(t) \leq \mathfrak{c}^2\Delta k^3. \qquad (**)$$

The Theorem then follows by application of Lemma 3.6.16 and Lemma 3.6.17. Starting with $(**)$ we use that $a^{it} = \cos(\log(a)t) + i\sin(\log(t))$ for $a \in \mathbb{R}_+$ and thus

$$|a^{it} - b^{it}| \leq |\cos(\log(a)t) - \cos(\log(b)t)| + |\sin(\log(a)t) - \sin(\log(b)t)| \leq 2t|\log(a) - \log(b)|,$$

for any $a, b \in \mathbb{R}_+$. This implies with **[A3]**

$$\frac{1}{2\pi}\int_{[-k,k]} |\mathbb{E}(V_0^{it} - \overline{V_1}^{it})|^2 d\lambda(t) \leq \frac{1}{2\pi}\int_{[-k,k]} t^2|\mathbb{E}(|\log(V_0) - \log(\overline{V}_1)|)|^2 d\lambda(t) \leq k^3\mathfrak{c}^2\Delta.$$

For $(*)$ let us denote by $\beta_{\overline{V}}(j) := \beta^{\text{mix}}(\sigma(\overline{V}_1), \sigma(\overline{V}_{j+1}))$. Then, since $\overline{V}_j = h_j((V_t)_{t\in((j-1)\Delta, j\Delta)})$ with h_j measurable,

$$\begin{aligned}\beta_{\overline{V}}(j) &\leq \beta^{\text{mix}}\left(\sigma(V_t : t \in (0,\Delta), \sigma(V_t : t \in (j\Delta, (j+1)\Delta)\right) \\ &\leq \beta^{\text{mix}}(\sigma(V_t : t \leq \Delta), \sigma(V_t : t \geq j\Delta),\end{aligned}$$

by monotonicity of β^{mix} with respect to nested σ-fields. Now for strictly stationary Markov processes, it is known that for any $s, \tau \in \mathbb{R}_+$,

$$\beta^{\text{mix}}(\sigma(V_t : t \le s), \sigma(V_t : t \ge s + \tau) = \beta^{\text{mix}}(V_0, V_\tau) = \beta(\tau), \tag{3.48}$$

compare Bradley (1986, Theorem 4.1), which implies with Lemma 3.6.6 that for all $t \in \mathbb{R}$

$$\frac{n}{4}\mathbb{V}\text{ar}(\widehat{\mathcal{M}}_{\overline{V}}(t)) \le \mathbb{E}(b(\overline{V}_1)) = \sum_{j=1}^{\infty} \beta_{\overline{V}}(j) \le \sum_{j=1}^{\infty} \beta(\Delta(j-1)) = \sum_{j=0}^{\infty} \beta(\Delta j).$$

Now exploiting that $t \mapsto \beta(t)$ is monotone decreasing, see (3.48), we bound the sum by the integral leading to

$$\frac{n}{4}\mathbb{V}\text{ar}(\widehat{\mathcal{M}}_{\overline{V}}(t)) \le 1 + \int_{\mathbb{R}_+} \beta(\Delta s) ds \le \Delta^{-1}\left(1 + \int_{\mathbb{R}_+} \beta(s) ds\right).$$

We deduce that $(*)$ holds true. □

Examples for volatility processes We will now study more precisely the assumptions **[A0]**-**[A3]** and propose examples of processes which fulfills these. Assumption **[A1]** can be checked for each example or by assuming conditions on the function b and a which are sufficient but not necessary, compare Comte and Genon-Catalot (2006). For **[A2]** we mainly refer to Pardoux and Veretennikov (2001). Based on Comte and Genon-Catalot (2006) assumption **[A3]** can be easily checked using the following Proposition.

Proposition 3.6.19 (Sufficient condition for **[A3]**):
If either

(i) $(V_t)_{t\ge 0}$ is a strictly stationary and ergodic diffusion process on $\mathbb{R}_+$ satisfying

$$dV_t = b(V_t)dt + a(V_t)dB_t,$$

and there exists $L > 0$ with

$$|a(x)| + |b(x)| \le L(1 + |x|)$$

for all $x \in \mathbb{R}$, with $b, a \in C^0(\mathbb{R}) \cap C^1(\mathbb{R}_+)$ and $\mathbb{E}(\sup_{t\in(0,\Delta)} V_t^{-2}), \mathbb{E}(V_0^2) < \infty$ hold true.

(ii) or $V_t = \exp(\mathcal{V}_t)$, where $(\mathcal{V}_t)_{t\ge 0}$ is a strictly stationary and ergodic diffusion process on $\mathbb{R}$ satisfying

$$d\mathcal{V}_t = \widetilde{b}(\mathcal{V}_t) + \widetilde{a}(\mathcal{V}_t)d\widetilde{B}_t,$$

and there exists $\widetilde{L} > 0$ with

$$|\widetilde{a}(x)| + |\widetilde{b}(x)| \le \widetilde{L}(1 + |x|)$$

for all $x \in \mathbb{R}$, $\widetilde{b}, \widetilde{a} \in C^1(\mathbb{R})$ and $\mathbb{E}(\mathcal{V}_0^2) < \infty$,

then $(V_t)_{t\ge 0}$ satisfy **[A3]**.

The proof of Proposition 3.6.19 can be found in the proof section. Let us collect some examples of processes fulfilling **[A0]**-**[A3]** which are drawn from Comte and Genon-Catalot (2006).

Example 3.6.20 (Exponential of an Ornstein-Uhlenbeck process)**:**
Let $(\mathcal{V}_t)_{t\geq 0}$ be the solution of the linear stochastic differential equation

$$d\mathcal{V}_t = -\alpha \mathcal{V}_t dt + \gamma dB_t, \quad \mathcal{V}_0 = \eta,$$

with $\alpha \in \mathbb{R}_+$,$\gamma \in \mathbb{R}$ and $\eta \sim \mathrm{N}_{(0,\rho^2)}$, where $\rho^2 = \gamma^2/(2\alpha)$. Then the unique solution $(\mathcal{V}_t)_{t\geq 0}$, where

$$\mathcal{V}_t = e^{-\alpha t}\eta + \gamma \int_0^t e^{-\alpha(t-s)} dB_s, \quad t \geq 0,$$

is a so-called Ornstein-Uhlenbeck process. Now, since $\eta \sim \mathrm{N}_{(0,\rho^2)}$, stochastically independent of $(B_t)_{t\geq 0}$, we deduce that $(V_t)_{t\geq 0}$ is strictly-stationary with $\mathcal{V}_t \sim \mathrm{N}_{(0,\rho^2)}$. Furthermore, it is a time-homogenous Markov process, ergodic, with continuous sample paths. This already implies that **[A1]** holds true for $V_t := \exp(\mathcal{V}_t)$. For **[A2]** we use Pardoux and Veretennikov (2001, Proposition 1) exploiting the fact, that the β-mixing coefficient can be expressed through the total variation between the stationary distribution of $(V_t)_{t\geq 0}$ and the distribution of V_t, if $V_0 = v_0 \in \mathbb{R}$ fixed. For the sake of simplicity we only mention this connection, while we refer Comte and Genon-Catalot (2006) for further reading. For **[A3]** we see that all assumption of Proposition 3.6.19 are fulfilled. The stationary density of V_0 is given by the density of a Log-Normal distribution

$$f_V(x) = \frac{1}{\sqrt{2\pi\rho^2 x}} \exp\left(-\frac{\log^2(x)}{2\rho^2}\right), \quad x \in \mathbb{R}_+,$$

where $f_V \in \mathbb{L}^2(\mathbb{R}_+, x)$ holds obviously true.

Let us quickly study the upcoming bias term $\|f_V - f_{V,k}\|_x^2$ in the case of an exponential Ornstein-Uhlenbeck process. Indeed, with Example 2.1.11 we get

$$\|f_V - f_{V,k}\|_x^2 = \frac{1}{\pi} \int_{(k,\infty)} \exp(-\rho^2 t^2) d\lambda(t) \leq \frac{\exp(-\frac{\rho^2 k^2}{2})}{\sqrt{2\pi}\rho}.$$

Then for the choice $k_o = \sqrt{\log(n)}\rho$ we get with Theorem 3.6.18

$$\mathbb{E}^n_{f_V}(\|f_V - \widehat{f}_{V,k_o}\|_x^2) \leq C(f_V, g)\left(n^{-1/2} + \Delta \log^{3/2}(n) + \frac{\log^{1/2}(n)}{n\Delta}\right)$$

where by setting $\Delta = \Delta_n := (n\log(n))^{-1/2}$ we obtain

$$\mathbb{E}^n_{f_V}(\|f_V - \widehat{f}_{V,k_0}\|_x^2) \leq C(f_V, g)\frac{\log(n)}{n^{1/2}}.$$

For a 1-dimensional Ornstein-Uhlenbeck process solving the differential equation

$$d\mathcal{V}_t = -\alpha V_t dt + \gamma dB_t, \quad \mathcal{V}_0 \sim \mathrm{N}_{(0,\rho^2)},$$

where $\rho^2 = \gamma^2/(2\alpha)$ we get by application of the Itô-formula, that the process $V_t := \mathcal{V}_t^2$ fulfills the stochastic differential equation

$$dV_t = (\gamma^2 - 2\alpha V_t)dt + 2\gamma\sqrt{V_t} dB_t, \quad V_0 \sim \Gamma_{(1/2,\gamma^2/(4\alpha))},$$

a special case of a so-called Cox-Ingersoll-Ross process which we will consider next.

Example 3.6.21 (Cox-Ingersoll-Ross process):
Let $(V_t)_{t\geq 0}$ be the solution of the stochastic differential equation

$$dV_t = \alpha(\beta - V_t)dt + \gamma\sqrt{V_t}dB_t, \quad V_0 = \eta,$$

where $\alpha, \beta, \gamma \in \mathbb{R}_+$m $\eta \sim \Gamma_{(\rho,\alpha/\gamma^2)}$, independent of $(B_t)_{t\geq 0}$ and $\rho = 2\alpha\beta/\gamma^2$. Since the drift function is linear and the volatility function is Hölder-continuous with $\beta = 1/2$, there exists a unique solution, the so-called Cox-Ingersoll-Ross process, or short CIR-process. This process is widely studied in the literature, for instance Cox et al. (1985). The process fulfills **[A1]**, if $\rho \geq 1$, compare Comte and Genon-Catalot (2006). For $\rho > 3$, we have $\mathbb{E}(\sup_{t\in(0,1)} V_t^{-2}) < \infty$, see Gloter (2000), which together with Proposition 3.6.19 implies **[A3]**. For the case $2\rho \in \mathbb{N}$ a solution of the SDE is given by

$$V_t = \sum_{j=1}^{2\rho} \mathcal{V}_{t,i}^2,$$

where $\mathcal{V}_t := (\mathcal{V}_{t,1}, \dots, \mathcal{V}_{t,2\rho})$ is a multivariate Ornstein-Uhlenbeck process with

$$d\mathcal{V}_t = -\frac{\alpha}{2}\mathcal{V}_t dt + \frac{\gamma}{2} d\boldsymbol{W}_t, \quad \mathcal{V}_0 \sim \mathrm{N}^{\otimes 2\rho}_{(0,\gamma^2/4\alpha)},$$

where $(\boldsymbol{W}_t)_{t\geq 0}$ is a 2ρ-dimensional Brownian motion. We can thus deduce that **[A2]** is in this case fulfilled. The stationary density of V_0 is given as the density of a Gamma distribution

$$f_V(x) = \frac{(2\alpha/\gamma^2)^\rho}{\Gamma(\rho)} x^{\rho-1} \exp\left(-\frac{2\alpha}{\gamma^2}x\right), \quad x \in \mathbb{R}_+.$$

Again, let us study the bias term for a Cox-Ingersoll-Ross process where we consider the case of $\alpha = \gamma^2/2$ and $\rho = 4$. Then we obtain with Example 2.1.7 and the Stirling formula

$$\|f_V - f_{V,k}\|_x^2 = \frac{1}{6\pi}\int_{(k,\infty)} |\Gamma(4+it)|^2 d\lambda(t) \leq C\int_{(k,\infty)} t^7 \exp(-\pi t) d\lambda(t) \leq C(f_V)\exp\left(-\frac{\pi}{2}k\right).$$

Choosing $k_o := 2\log(n)/3\pi$ and $\Delta = (n\log(n))^{-1/2}$ we get

$$\mathbb{E}^n_{f_V}(\|f_V - \widehat{f}_{\Delta,k_o}\|_x^2) \leq C(f_V, g)\frac{1}{n^{1/3}}.$$

Example 3.6.22 (Exponential of a CIR-process):
Let $\rho \in \mathbb{N}$ and $\lambda \in \mathbb{R}_+$. Then $V_t = \exp(\mathcal{V}_t)$ with

$$d\mathcal{V}_t = (\frac{2\rho}{\lambda} - \mathcal{V}_t)dt + \sqrt{\frac{2}{\lambda}}\sqrt{\mathcal{V}_t}dB_t, \quad \mathcal{V}_0 \sim \Gamma_{(\rho,\lambda)}$$

where η independent of $(B_t)_{t\geq 0}$. $(V_t)_{t\geq 0}$ fulfills **[A1]** since $(\mathcal{V}_t)_{t\geq 0}$ fulfills **[A1]**. Similarly, $(V_t)_{t\geq 0}$ is β-mixing with integrable coefficients as $(\mathcal{V}_t)_{t\geq 0}$, implying that **[A2]** holds true. To see **[A3]** we again use Proposition 3.6.19, where we should stress out in Proposition 3.6.19 the assumption on $\widetilde{a}, \widetilde{b}$ can be relaxed to $\widetilde{a}, \widetilde{b} \in C^0(\mathbb{R}) \cap C^1(\mathbb{R}_+)$ for strict positive processes and we use the expansion $\widetilde{a}(x) := \lambda^{-1/2}\mathbb{1}_{\mathbb{R}_+}(x)\sqrt{x}$. The stationary density of V_0 is given by the density of a Log-Gamma distribution

$$f_V(x) = \frac{\lambda^\rho}{\Gamma(\rho)} x^{-\lambda-1}\log(x)^{\rho-1}\mathbb{1}_{(1,\infty)}(x), \quad x \in \mathbb{R}_+.$$

For the squared bias of an exponential of a Cox-Ingersoll-Ross process with $\lambda \in \mathbb{R}_+$ and $\rho \in \mathbb{N}$, we have with Example 2.1.10

$$\|f_V - f_{V,k}\|_x^2 = \frac{\lambda^2 2\rho}{\pi} \int_{(k,\infty)} (\lambda^2 + t^2)^{-\rho} d\lambda(t) \le C(\lambda, \rho) k^{-2(\rho-1/2)}.$$

Choosing $k_o := \log(n)/2\pi$ and $\Delta = (n \log(n))^{-1/2}$ we get

$$\mathbb{E}^n_{f_V}(\|f_V - \widehat{f}_{\Delta,k_o}\|_x^2) \le C(f_V, g) \frac{1}{\log(n)^{2(\rho-1/2)}}.$$

Data-driven choice of $k \in \mathbb{R}_+$ First we reduce the space of possible cut-off parameters to

$$\mathcal{K}_n := \{k \in [\![\log(n)/\pi]\!] \;:\; \Lambda_g(k) \le n\}$$

based on Section 3.3.3, since we are in the case of a super smooth error density, compare with the discussion under Lemma 3.6.16. As seen before, this reduction is rather natural. Then we define the model selection method for $\chi > 0$

$$\widehat{k} := \arg\min_{k \in \mathcal{K}_n} \; -\|\widehat{f}_{\Delta,k}\|_x^2 + \mathrm{pen}(k), \quad \mathrm{pen}(k) := \chi k \Lambda_g(k) n^{-1}. \tag{3.49}$$

Theorem 3.6.23 (Data-driven choice of $k \in \mathbb{R}_+$)**:**
Let $f_V \in \mathbb{L}^2(\mathbb{R}_+, x)$ and **[A0]-[A3]** hold true. Then there exists $\chi_0 \in \mathbb{R}_+$, such that for all $\chi \ge \chi_0$

$$\mathbb{E}(\|\widehat{f}_{\Delta,\widehat{k}} - f_V\|_x^2) \le C \inf_{k \in \mathcal{K}_n} \left(\|f_V - f_{V,k}\|_x^2 + \mathrm{pen}(k)\right) + C\Delta \log^3(n) + \frac{C(g)}{n}, + \frac{C(\beta)\log(n)}{n\Delta}$$

where C is a numerical constant and $C(g)$, respectively $C(\beta)$, are positive constant only depending on g, respectively on $\int_{\mathbb{R}_+} \beta(t)dt$.

Proof of Theorem 3.6.23. As $\mathbb{E}(\|f_V - \widehat{f}_{\Delta,\widehat{k}}\|_x^2) = \|f_V - f_{V,K_n}\|_x^2 + \mathbb{E}(\|f_{V,K_n} - \widehat{f}_{\Delta,\widehat{k}}\|_x^2)$, it follows

$$\mathbb{E}(\|f_V - \widehat{f}_{\Delta,\widehat{k}}\|_x^2) \le \|f_V - f_{V,K_n}\|_x^2 + 2\|f_{V,K_n} - f_{\Delta,K_n}\|_x^2 + 2\mathbb{E}(\|f_{\Delta,K_n} - \widehat{f}_{\Delta,\widehat{k}}\|_x^2).$$

For the last summand we can apply the result of Theorem 3.6.9 with $f_{K_n} := f_{\Delta,K_n}$ and $\widehat{f}_k := \widehat{f}_{\Delta,k}$ to get

$$\begin{aligned}\mathbb{E}(\|f_{\Delta,K_n} - \widehat{f}_{\Delta,\widehat{k}}\|_x^2) \le 3 \inf_{k \in \mathcal{K}_n} \left(\|f_{\Delta,K_n} - f_{\Delta,k}\|_x^2 + \mathrm{pen}(k)\right) + \frac{C(g)}{n} \\ + \int_{[-K_n,K_n]} \mathrm{Var}^n_f(\widehat{\mathcal{M}}_{\overline{V}}(t)) d\lambda(t).\end{aligned}$$

Here, several summands can be omitted since the penalty term $(\mathrm{pen}(k))_{k\in\mathcal{K}n}$ has not to be estimated in the case of $c = 1$. Now since

$$\begin{aligned}\|f_{\Delta,K_n} - f_{\Delta,k}\|_x^2 &\le 3(\|f_{\Delta,K_n} - f_{V,K_n}\|_x^2 + \|f_{V,K_n} - f_{V,k}\|_x^2 + \|f_{V,k} - f_{\Delta,k}\|_x^2) \\ &\le 6\|f_{\Delta,K_n} - f_{V,K_n}\|_x^2 + 3\|f_{V,K_n} - f_{V,k}\|_x^2,\end{aligned}$$

we get

$$\mathbb{E}(\|f_V - \widehat{f}_{\Delta,\widehat{k}}\|_x^2) \leq C \left(\inf_{k\in\mathcal{K}_n} \left(\|f_V - f_{V,k}\|_x^2 + \mathrm{pen}(k)\right) + \|f_{\Delta,K_n} - f_{V,K_n}\|_x^2 \right) + \frac{C(g)}{n}$$
$$+ 8 \int_{[-K_n,K_n]} \mathbb{V}\mathrm{ar}_f^n(\widehat{\mathcal{M}}_{\overline{V}}(t)) d\lambda(t)$$
$$\leq C \inf_{k\in\mathcal{K}_n} \left(\|f_V - f_{V,k}\|_x^2 + \mathrm{pen}()k)\right) + C\Delta K_n^3 + \frac{C(\beta)K_n}{n\Delta}$$

following the proof steps of Theorem 3.6.18. Now $K_n \leq \log(n)/(2\pi)$ implies the claim. □

For the estimation of the stationary density in the stochastic volatility model, we derived an estimator based on the Mellin transform for the choice $c = 1$. Here, we have seen that this model is similar to a multiplicative deconvolution model with super smooth error density and strictly stationary process. Based on the results of Section 3.3 and Section 3.6 it can be seen that the developed results for the case $c = 1$, Theorem 3.6.18 and Theorem 3.6.23, can be expand to any case $c \in \mathbb{R}$, which fulfills the necessary moment conditions. More precisely, regarding Proposition 3.6.4, we have to assume that $\mathbb{E}_g(U_1^{2(c-1)}) < \infty$, which is only fulfilled for $U \sim \Gamma_{(1/2,1/2)}$ if $c > 3/4$. For the adaptive estimator, Theorem 3.6.9, we have to assume that $\mathbb{E}_g(U_1^{4(c-1)}) < \infty$, which holds only true for $c > 7/8$.

Example 3.6.24 (Adaptive rates for Examples 3.6.20-3.6.22)**:**
With the choice of $\Delta = (n\log^2(n))^{-1/2}$ we get

$$\mathbb{E}_{f_V}^n(\|f_V - \widehat{f}_{\Delta,\widehat{k}}\|_x^2) \leq C(f_V, g) \cdot \begin{cases} \frac{\log^2(n)}{n^{1/2}}, & \text{Example 3.6.20,} \\ \frac{1}{n^{1/3}} & \text{Example 3.6.21,} \\ \log(n)^{-2(\rho-1/2)}, & \text{Example 3.6.22.} \end{cases}$$

Here, it is worth pointing out that Δ is not dependent on the processes and that only in the case of Example 3.6.20 a slightly slower rate is achievable. This is due to the fact, that the additional bias $\|f_{V,k} - f_{\Delta,k}\|_x^2$ is evaluated in the maximal dimension $K_n = [\![\log(n)/\pi]\!]$ and not in the oracle dimension $k_o = \sqrt{\log(n)}\rho$, since the latter is unknown. Nevertheless, the maximal dimension K_n is a known, sophisticated upper bound for all possible oracle choices of k_o, which are unknown and may depend on f_V.

3.6.5 Numerical results

Considering the Examples 3.6.20-3.6.22 the processes can be expressed as measurable transformations of Ornstein-Uhlenbeck processes. Thus sampling from these processes reduces itself to sampling from Ornstein-Uhlenbeck processes, which we will discuss first. Let $\sigma^2 > 0$ be the desired variance, then the Ornstein-Uhlenbeck process $(O_t)_{t\geq 0}$ solves the stochastic differential equation

$$dO_t = -\frac{1}{2}O_t dt + \sigma dB_t, \quad O_0 \sim \mathrm{N}_{(0,\sigma^2)},$$

where O_0 and $(B_t)_{t\geq 0}$ are independent. The explicit strong solution is then given by

$$O_t = e^{-\frac{t}{2}} O_o + \sigma e^{-\frac{t}{2}} \int_0^t e^{\alpha s} dB_s, \quad t \geq 0,$$

compare Example 3.6.20. A discrete sample scheme is then given for $\Delta \in (0,1)$ by exchanging the upcoming stochastic integral by a discretization characterized by the step size $\delta \in (0,1)$, w.l.o.g. $\delta^{-1} \in \mathbb{N}$, given by $(\widehat{O}_{j\delta\Delta})_{j\in\mathbb{N}}$, where

$$\begin{aligned}\widehat{O}_{j\delta\Delta} :&= e^{-\frac{\delta\Delta j}{2}} O_o + \sigma e^{-\frac{\delta\Delta j}{2}} \sum_{i=1}^{j} e^{(i-1)\delta\Delta}(B_{i\delta\Delta} - B_{(i-1)\delta\Delta})\\ &= e^{-\frac{\delta\Delta}{2}} \widehat{O}_{(j-1)\Delta\delta} + \sigma e^{-\frac{\Delta\delta}{2}} (W_{j\Delta\delta} - W_{(j-1)\Delta\delta}).\end{aligned}$$

The process $(\widetilde{O}_{j\delta\Delta})_{j\in\mathbb{N}}$ has the form of an autoregressive process, since the increments of the Brownian motion are identically and independently distributed. It is clear, that the process $(\widehat{O}_{j\delta\Delta}))_{j\in\mathbb{N}}$ is not strictly stationary, since the initial distribution of $O_o \sim \mathrm{N}_{(0,\sigma^2)}$ differs from the invariant distribution $\mathrm{N}_{(0,\sigma^2 \frac{\delta\Delta}{\exp(\delta\Delta)-1})}$. As we can see, the approximation leads to a digression of the variance of the invariant distribution. Nevertheless, let w.l.o.g. $\Delta = 1$. Then, for $\delta \to 0$ we get

$$\lim_{\delta\to 0} \frac{\delta}{\exp(\delta)-1} = 1, \quad \text{more precisely for all} \quad \delta > 0 \left| \frac{\delta}{\exp(\delta)-1} - 1 \right| \leq (e-1)\delta$$

using the rule of l'Hôpital, respectively the infinite sum representation of the exponential function. Throughout the simulations, we consider two different regimes.

[R1] Stationary sample: We estimate the density f_X of a stationary sequence $(X_j)_{j\in[\![n]\!]}$, β-mixing, given the sample

$$Y_j = X_j U_j, \quad j \in [\![n]\!],$$

with $(U_j)_{j\in[\![n]\!]}$ i.i.d. sequence independent of $(X_j)_{j\in[\![n]\!]}$.

[R2] Stochastic volatility model: We estimate the density f_V of a stationary volatility process $(V_t)_{t\geq 0}$ given the sample

$$Y_j = \overline{V}_j U_j, \quad \overline{V}_j := \Delta^{-1} \int_{(j-1)\Delta}^{j\Delta} V_s ds, \quad j \in [\![n]\!], \Delta \in (0,1),$$

where $(U_j)_{j\in[\![n]\!]}$ is i.i.d and stochastically independent of $(V_t)_{t\geq 0}$.

For the generation of a sample in the regime **[R1]**, we make use of the fact that $(X_t)_{t\in\mathbb{N}}$, the discrete-time sample of $(X_t)_{t\geq 0}$ of a strictly-stationary, β-mixing process, is again strictly-stationary and β-mixing, which is a direct consequence of the definition of both concepts.
Thus, choosing $\Delta = 1$ and $\delta^{-1} \in \mathbb{N}$ we generate the sample $(\widehat{X}_{\ell\delta})_{\ell\in\mathbb{N}}$ as described before by a measurable transformation of approximations of Ornstein-Uhlenbeck processes and use the subsample $X_j := \widehat{X}_j = \widehat{X}_{(j\delta^{-1})\delta}$, which correspond to taking $\ell = j\delta^{-1} \in \mathbb{N}$. Analogously, we can define samples $(X_j)_{j\in[\![n]\!]}$ for arbitrary values of $\Delta \in (0,1)$, in other words, with varying distance between the discrete-time samples.
For the regime **[R2]** we generate first for $\delta^{-1} \in \mathbb{N}$ and $\Delta \in (0,1)$ the sample $(\widehat{V}_{j\delta\Delta})_{j\in\mathbb{N}}$ as described by a measurable transformation of approximation of Ornstein-Uhlenbeck processes. In the second step, we use this sample to sample from $\overline{V}_j$ using a Riemann-sum approximation

$$\widehat{\overline{V}}_j := \Delta^{-1} \sum_{i=1}^{\delta^{-1}} \delta\Delta \widehat{V}_{(j-1)\Delta + i\delta\Delta}, \quad j \in \mathbb{N}.$$

For sake of simplicity, we will write X_j instead of $\widehat{X}_j$, respectively $\overline{V}_j$ instead of $\widehat{\overline{V}}_j$, keeping in mind that we introduced a numerical error due to the numerical approximations in dependence of $\delta \in (0,1)$, which is vanishing for $\delta \to 0$.

Densities and error densities For the illustration of the behavior of the fully data-driven estimator $\widehat{f}_{\widehat{k}}$, respectively $\widehat{f}_{\widehat{k},\Delta_n}$, we consider the following examples of densities f, respectively volatility densities f_V.

(i) Log-Gamma Distribution: $f(x) = 5^5\Gamma(5)^{-1}x^{-6}\log(x)^4\mathbb{1}_{(1,\infty)}(x), x \in \mathbb{R}_+$,

(ii) Gamma Distribution: $f(x) = \Gamma(5)^{-1}x^4\exp(-x)\mathbb{1}_{\mathbb{R}_+}(x), x \in \mathbb{R}_+$, and

(iii) Log-Normal Distribution: $f(x) = (0.32\pi x^2)^{-1/2}\exp(-\log^2(x)/0.32)\mathbb{1}_{(0,\infty)}(x), x \in \mathbb{R}_+$.

These densities have been already considered in Section 3.1, Section 3.2 and are at the same time the stationary densities of the Examples 3.6.20-3.6.22. For the error densities, we consider a smooth and a super smooth error density, given by

$a)$ Smooth error: $g(x) := \mathbb{1}_{(1/2,3/2)}(x), x \in \mathbb{R}_+$ and

$b)$ Super smooth error: $g(x) := (2\pi)^{-1/2}x^{-1/2}\exp(-x/2)\mathbb{1}_{\mathbb{R}_+}(x), x \in \mathbb{R}_+$.

The error density $a)$ is a uniform distribution on the interval $(1/2, 3/2)$ which is a smooth error density with $\gamma = 1$, while $b)$ is the density of a χ^2 distribution which is a super smooth error density naturally appearing in the stochastic volatility model.
Although we have not proposed a data-driven choice of the parameter $k \in \mathbb{R}_+$ in the context of smooth error densities and a stationary sample, we use this additionally case as an illustration the effect of the dependency in comparison to the effect of the super smooth error density. In the smooth case, we use the data-driven selection defined in (3.15).

Stationary sample [R1] We consider the estimation for $c = 1$ and for the error density $a)$, where we use the value of $\chi = 7$ as proposed in Section 3.2.5 while for $b)$ we choose $\chi = 1.7$ based on a preliminary simulation study. Throughout this paragraph we chosen $\delta := 0.1/\Delta$ to ensure that, even if we vary Δ, the underlying stationary distribution of our data stays the same to ensure comparability between the plots. For Figure 3.17 and Figure 3.19 we have $\Delta = 1$ and $\delta = 0.1$. In Figure 3.18 we let Δ, the step length between the discrete-time samples $(X_j)_{j\in[\![n]\!]}$ vary, $\Delta \in \{0.2, 0.5, 1\}$, to study the influence of the dependency structure on the behavior of our estimator.

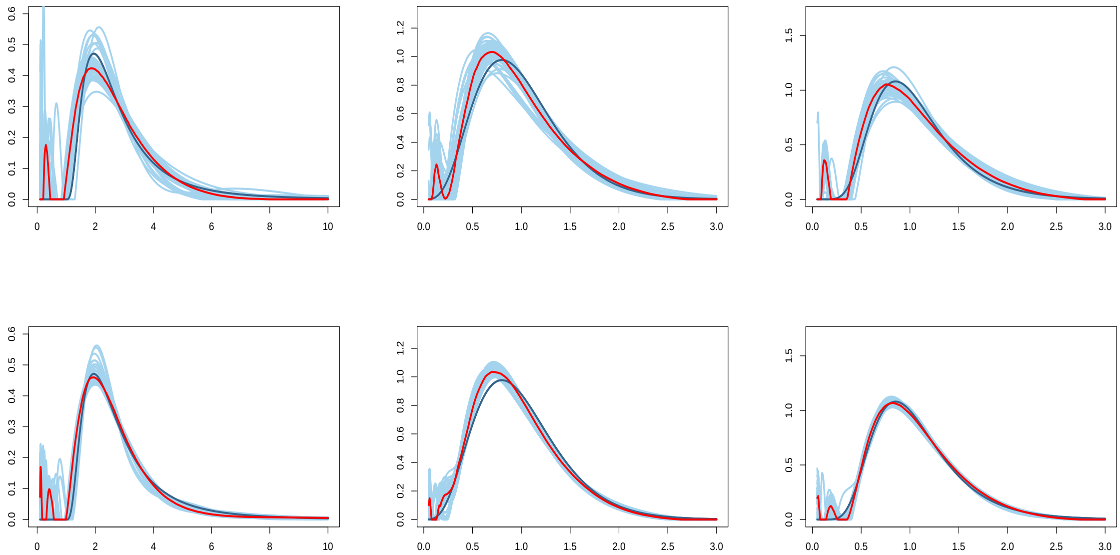

Figure 3.17: Considering the estimators $\widehat{f}_{\widehat{k}}$, we depict 50 Monte-Carlo simulations with varying stationary density (i) (left), (ii) (middle) and (iii) (right) for $n = 500$ (top) and $n = 5000$ (bottom), where $\delta = 0.1, \Delta = 1$, using the error density $a)$. The true density function f is given by the black curve, while the red curve is the point-wise empirical median of the 50 estimates.

Figure 3.17 visualizes the performance of our estimator for a stationary sample $(X_j)_{j\in[\![n]\!]}$ under varying sample size for different choices of stationary densities.
As the Ornstein-Uhlenbeck process is β-mixing with integrable coefficients, so are the discrete time samples of the exponential of an Ornstein-Uhlenbeck process, the CIR-process and the exponential of an CIR-process, see Example 3.6.20, Example 3.6.21 and Example 3.6.22. More precisely, it can be seen that the parameter Δ, corresponding to the length of the distance between the discrete observation times $\Delta = t_i - t_{i-1}$ of $(X_{t_i})_{i\in[\![n]\!]}$, effects the dependency between the observations, that is for $\Delta \to 0$ the dependency is increasing. This effect can be seen in the following Figure 3.18.

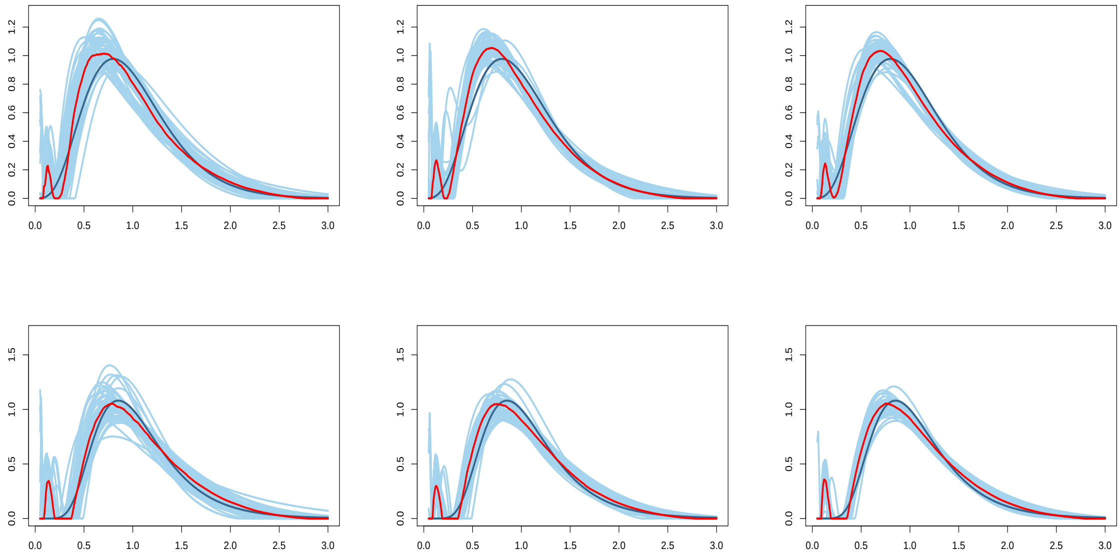

Figure 3.18: Considering the estimators $\widehat{f}_{\widehat{k}}$, we depict 50 Monte-Carlo simulations with stationary density (ii) (top) and (iii) (bottom) for $n = 500$ where $\Delta = 0.2$ (left), $\Delta = 0.5$ (middle) and $\Delta = 1$ (right) using the error density $a)$. The true density function f is given by the black curve while the red curve is the point-wise empirical median of the 50 estimates.

For the Figure 3.18 we considered different choices of Δ and changed $\delta \in (0, 1)$, such that $\delta * \Delta = 0.1$ to ensure that we only observe a difference in the underlying dependency. For $\Delta \to \infty$, the dependency between the observation vanishes.
In comparison to Figure 3.17, we see in Figure 3.19 that even for higher sample size the estimator $\widehat{f}_{\widehat{k}}$ has still a visible bias. Indeed, from a theoretical point of view, the different error densities have an enormous effect on the behavior of the variance term. More precisely, for the chi-squared distributed error, case $b)$, the variance term is increasing exponentially. Thus large values of $k \in \mathbb{R}_+$ can almost not be chosen. For the simulation with error density $b)$ we used a finer grid than $\mathcal{K}_n$ as dimensions higher then $k = 3$ have never been chosen by our data-driven method.

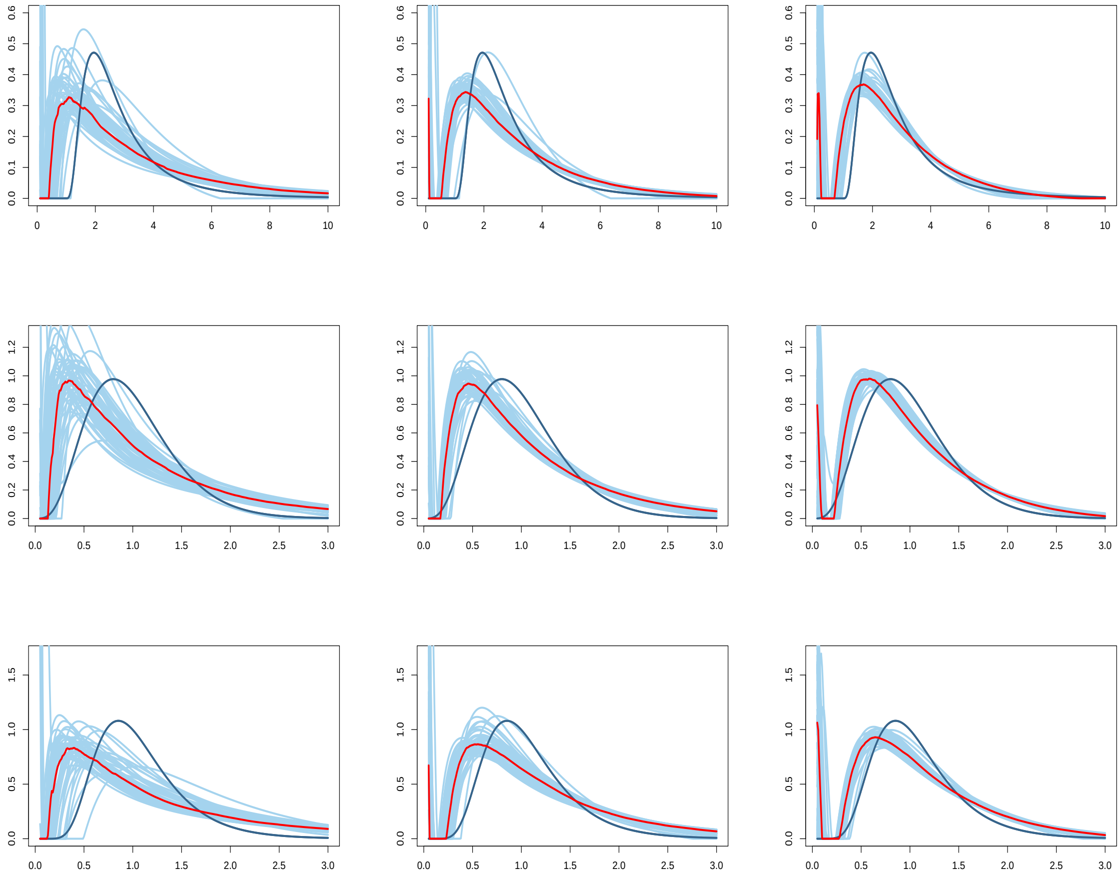

Figure 3.19: Considering the estimators $\widehat{f}_{\widehat{k}}$, we depict 50 Monte-Carlo simulations with varying sstationary density (i) (top), (ii) (middle) and (iii) (bottom) for $n = 5 * 10^2$ (left), $n = 5 * 10^3$ (middle) and $n = 5 * 10^4$ (right), where $\delta = 0.1, \Delta = 1$ and for the error density b). The true density function f is given by the black curve while the red curve is the point-wise empirical median of the 50 estimates.

Stochastic volatility models [R2] To illustrate the behavior of our estimator in the stochastic volatility we focus on the density (iii), the Log-Normal distribution. We will study two different settings to visualize the theoretical results. First, we study the same estimator $\widehat{f}_{\widehat{k}}$ based on different samples, namely $(V_{j\Delta})_{j\in[\![n]\!]}$ and $(\overline{V}_j)_{j\in[\![n]\!]}$ with $\Delta = 0.2$. As we do not consider measurement errors in this setting, we use the direct estimator by setting $\mathcal{M}_1[g] = 1$.
In the second step, we will analyze the effects of the error density. For the stochastic volatility model, we have seen that the increments can be written as observations of a multiplicative measurement errors model with $\Gamma_{(1/2,1/2)}$-distributed errors, namely case (b). To visualize now the impact, we consider the direct model based on the observations $(\overline{V}_j)_{j\in[\![n]\!]}$ and compare it with the error model observations $(Y_j)_{j\in[\![n]\!]}$, $Y_j = \overline{V}_j U_j$ where $U_j \sim \Gamma_{(1/2,1/2)}$.
For $\chi > 0$ we choose $\chi = 42.27$ for the direct observations and $\chi = 1.7$ for the stochastic volatility model observations based on preliminary simulations.

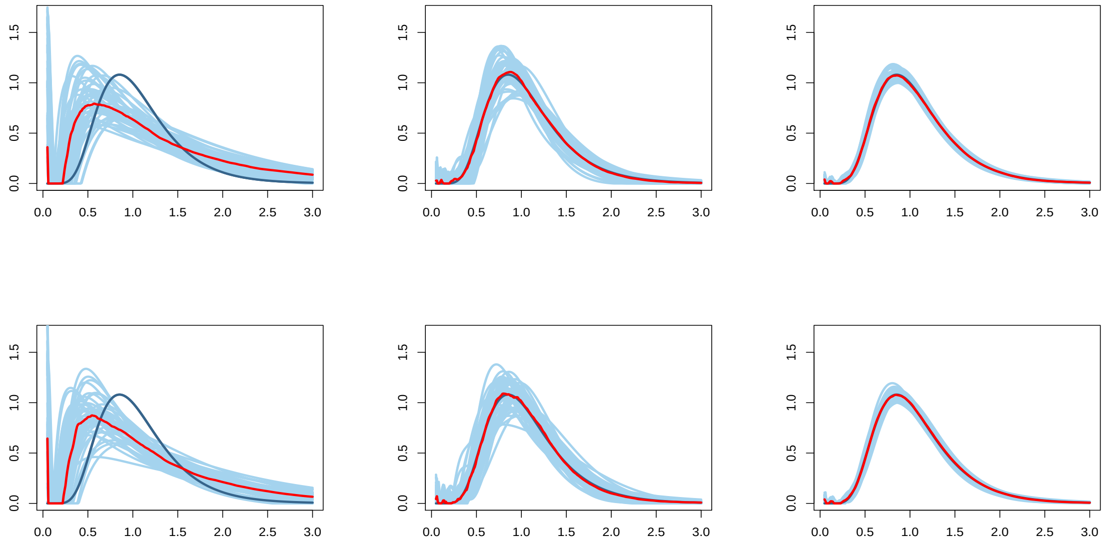

Figure 3.20: Considering the estimators $\widehat{f}_{\widehat{k}}$, we depict 50 Monte-Carlo simulations with stationary density (iii) for $n = 5*10^2$ (left), $n = 5*10^3$ (middle) and $n = 5*10^4$ where $\Delta = 0.2$ without noise based on the sample $(V_{j\Delta})_{j\in[\![n]\!]}$ (top) and $(\overline{V}_j)_{j\in[\![n]\!]}$ (bottom). The true density function f is given by the black curve while the red curve is the point-wise empirical median of the 50 estimates.

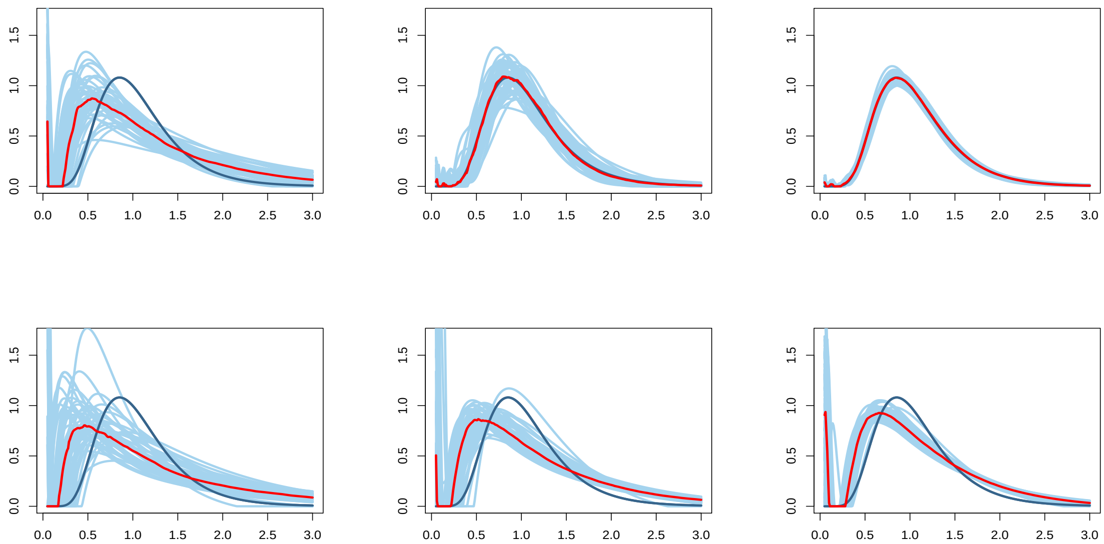

Figure 3.21: Considering the estimators $\widehat{f}_{\widehat{k}}$, we depict 50 Monte-Carlo simulations with stationary density (iii) with varying sample size $n = 5*10^2$ (left), $n = 5*10^3$ (middle) and $n = 5*10^4$ where $\Delta = 0.2$ based on the sample $(\overline{V}_j)_{j\in[\![n]\!]}$ (top) and $(Y_j)_{j\in[\![n]\!]}$ (bottom). The true density function f is given by the black curve while the red curve is the point-wise empirical median of the 50 estimates.

In Figure 3.20, we see that the difference between the samples $(V_{j\Delta})_{j\in[\![n]\!]}$ and $(\overline{V}_j)_{j\in[\![n]\!]}$ is rather minor for $\Delta = 0.2$ and the Log-Normal distribution.
Similar effects as in Figure 3.18 and Figure 3.19 can be seen in Figure 3.21. As implied by the theoretical results the underlying inverse problem with super smooth error density mainly characterize the risk of the estimator.

3.6.6 Conclusion

In this section, we studied the fully data-driven estimator $\widehat{f}_{\widehat{k}}$ proposed in Section 3.2 and Section 3.3 in the dependent case, twe considered that the sample $(X_j)_{j\in[\![n]\!]}$ is strictly stationary but not independent. We were able to show similar results to the independent case up to an additional term characterized by the underlying dependency. Furthermore, we considered the model of stationary density estimation in a stochastic volatility model and identified it as an multiplicative measurement error model with super smooth error and stationary sample. Thus previous shown results were used to construct and estimator of the volatility density. With the help of a Monte-Carlo simulation we have shown the reasonable behavior of our estimator for finite sample.

3.6.7 Proofs

Proof of Lemma 3.6.11. Let $k \in \mathcal{K}_n$ and let us keep in mind that $[-k', k'] = \mathrm{supp}(\mathcal{M}_{\boldsymbol{c}}[f_{k'}])$, for $k' \in \mathcal{K}_n$. By definition of $\mathcal{K}_n$, we have for all $k' \in \mathcal{K}_n$ holds $[-k', k] \subseteq [-K_n, K_n]$. Further, we have for any $k' \in \mathcal{K}_n$ that $\|\widehat{f}_{K_n}\|^2_{x^{2c-1}} - \|\widehat{f}_{k'}\|^2_{x^{2c-1}} = \|\widehat{f}_{K_n} - \widehat{f}_{k'}\|^2_{x^{2c-1}}$, implying with (3.43)

$$\|\widehat{f}_{\widehat{k}} - \widehat{f}_{K_n}\|^2_{x^{2c-1}} + \widehat{\mathrm{pen}}(\widehat{k}) \leq \|\widehat{f}_k - \widehat{f}_{K_n}\|^2_{x^{2c-1}} + \widehat{\mathrm{pen}}(k).$$

Now for every $k' \in \mathcal{K}_n$ we have

$$\|\widehat{f}_{k'} - f_{K_n}\|^2_{x^{2c-1}} = \|\widehat{f}_{k'} - \widehat{f}_{K_n}\|^2_{x^{2c-1}} + \|\widehat{f}_{K_n} - f_{K_n}\|^2_{x^{2c-1}} + 2\langle \widehat{f}_{k'} - \widehat{f}_{K_n}, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}},$$

which implies that

$$\begin{aligned}\|\widehat{f}_{\widehat{k}} - f_{K_n}\|^2_{x^{2c-1}} - \|\widehat{f}_k - f_{K_n}\|^2_{x^{2c-1}} &= \|\widehat{f}_{\widehat{k}} - \widehat{f}_{K_n}\|^2_{x^{2c-1}} - \|\widehat{f}_k - \widehat{f}_{K_n}\|^2_{x^{2c-1}} + 2\langle \widehat{f}_{\widehat{k}} - \widehat{f}_k, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}}\\ &\leq \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k}) + 2\langle \widehat{f}_{\widehat{k}} - \widehat{f}_k, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}}. \end{aligned} \tag{3.50}$$

Further, we see

$$\begin{aligned}\langle \widehat{f}_{\widehat{k}} - \widehat{f}_k, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}} &= \langle (\widehat{f}_{\widehat{k}} - f_{\widehat{k}}) + (f_{\widehat{k}} - f_k) + (f_k - \widehat{f}_k), \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}}\\ &= \|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}} + \langle f_{\widehat{k}} - f_k, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}} - \|\widehat{f}_k - f_k\|^2_{x^{2c-1}}. \end{aligned} \tag{3.51}$$

Now combining (3.50) and (3.51), we get

$$\begin{aligned}\|\widehat{f}_{\widehat{k}} - f_{K_n}\|^2_{x^{2c-1}} \leq \|f_k - f_{K_n}\|^2_{x^{2c-1}} - \|\widehat{f}_k - f_k\|^2_{x^{2c-1}} + 2\langle f_{\widehat{k}} - f_k, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}}\\ + \widehat{\mathrm{pen}}(k) + 2\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}} - \widehat{\mathrm{pen}}(\widehat{k}). \end{aligned} \tag{3.52}$$

Let us consider the term $|2\langle f_{\widehat{k}} - f_k, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}}|$. First we remind that for any $k' \in \mathcal{K}_n$

$$\|\widehat{f}_{k'} - f_{k'}\|^2_{x^{2c-1}} = \frac{1}{2\pi}\int_{\mathbb{R}} \mathbb{1}_{[-k',k']}(t)\frac{|\mathcal{M}_c[f_Y](t) - \widehat{\mathcal{M}}_c(t)|^2}{|\mathcal{M}_c[g](t)|^2}d\lambda(t).$$

Setting $A^* := [-\widehat{k}, \widehat{k}] \cup [-k, k]$ we have $\mathcal{M}_c[f_{\widehat{k}} - f_k] = \mathcal{M}_c[f](\mathbb{1}_{[-\widehat{k},\widehat{k}]} - \mathbb{1}_{[-k,k]})$, implying that $\operatorname{supp}(\mathcal{M}_c[f_{\widehat{k}} - f_k]) \subseteq A^* \subseteq [-K_n, K_n]$ by definition of K_n. Using the Cauchy-Schwarz inequality and $2ab \leq a^2 + b^2$, we deduce

$$\begin{aligned}
|2\langle f_{\widehat{k}} - f_k, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}}| &= \frac{2}{(2\pi)^2}\left|\int_{A^*} \mathcal{M}_c[f_{\widehat{k}} - f_k](t)\frac{\widehat{\mathcal{M}}_c(-t) - \mathcal{M}_c[f_Y](-t)}{\mathcal{M}_c[g](-t)} d\lambda(t)\right| \\
&\leq \frac{1}{4}\|f_{\widehat{k}} - f_k\|^2_{x^{2c-1}} + \frac{4}{2\pi}\int_{\mathbb{R}} \mathbb{1}_{A^*}(t)\frac{|\mathcal{M}_c[f_Y](t) - \widehat{\mathcal{M}}_c(t)|^2}{|\mathcal{M}_c[g](t)|^2} d\lambda(t) \\
&\leq \frac{1}{2}\|f_{\widehat{k}} - f_{K_n}\|^2_{x^{2c-1}} + \frac{1}{2}\|f_k - f_{K_n}\|_{x^{2c-1}} + 4\|\widehat{f}_k - f_k\|^2_{x^{2c-1}} + 4\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}}
\end{aligned}$$

using that $\mathbb{1}_{A^*} \leq \mathbb{1}_{A_k} + \mathbb{1}_{A_{\widehat{k}}}$. Thus, we get

$$\begin{aligned}
&|2\langle f_{\widehat{k}} - f_k, \widehat{f}_{K_n} - f_{K_n}\rangle_{x^{2c-1}}| \\
&\quad \leq \frac{1}{2}\|f_{K_n} - f_k\|^2_{x^{2c-1}} + \frac{1}{2}\|\widehat{f}_{\widehat{k}} - f_{K_n}\|^2_{x^{2c-1}} + 4\|\widehat{f}_k - f_k\|^2_{x^{2c-1}} + \frac{7}{2}\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}}
\end{aligned}$$

which implies with (3.52)

$$\|\widehat{f}_{\widehat{k}} - f_{K_n}\|^2_{x^{2c-1}} \leq 3\|f_k - f_{K_n}\|^2_{x^{2c-1}} + 6\|\widehat{f}_k - f_k\|^2_{x^{2c-1}} + 2\widehat{\operatorname{pen}}(k) + 11\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}} - 2\widehat{\operatorname{pen}}(\widehat{k}).$$

Since $\mathbb{E}^n_{f_Y}(\widehat{\operatorname{pen}}(k)) = \operatorname{pen}(k)$ and for $\chi_0 \geq 6$

$$\begin{aligned}
6\mathbb{E}^n_{f_Y}(\|\widehat{f}_k - f_k\|^2_{x^{2c-1}}) &\leq \frac{6\sigma_Y \Delta_g(k)}{n} + \frac{6}{2\pi}\int_{[-k,k]} \mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t) d\lambda(t) \\
&\leq \operatorname{pen}(k) + \int_{[-k,k]} \mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t)) d\lambda(t),
\end{aligned}$$

we get

$$\begin{aligned}
\mathbb{E}^n_{f_Y}(\|\widehat{f}_{\widehat{k}} - f_{K_n}\|^2_{x^{2c-1}}) \leq & 3\left(\|f_{K_n} - f_k\|^2_{x^{2c-1}} + \operatorname{pen}(k)\right) + 11\mathbb{E}^n_{f_Y}\left(\|\widehat{f}_{\widehat{k}} - f_{\widehat{k}}\|^2_{x^{2c-1}} - \frac{1}{12}\operatorname{pen}(\widehat{k})\right)_+ \\
&+ \mathbb{E}^n_{f_Y}((\operatorname{pen}(\widehat{k}) - 2\widehat{\operatorname{pen}}(\widehat{k}))_+) + \int_{[-K_n,K_n]} \mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t)) d\lambda(t).
\end{aligned}$$

□

Proof of Proposition 3.6.19. Let us begin with (ii). We have

$$\begin{aligned}
\log(\overline{V}_1) - \log(V_0) &= \log\left(\Delta^{-1}\int_0^\Delta e^{\mathcal{V}_t - \mathcal{V}_0} dt\right) \leq \sup_{t\in(0,\Delta)} \mathcal{V}_t - \mathcal{V}_0 \text{ and} \\
\log(V_0) - \log(\overline{V}_1) &= -\log\left(\Delta^{-1}\int_0^\Delta e^{\mathcal{V}_t - \mathcal{V}_0} dt\right) \leq -\inf_{t\in(0,\Delta)} \mathcal{V}_t - \mathcal{V}_0
\end{aligned}$$

implying that $\mathbb{E}(|\log(\overline{V}_1) - \log(V_0)|) \leq \mathbb{E}(\sup_{t\in[0,\Delta]} |\mathcal{V}_t - \mathcal{V}_0|)$. Next since $\mathcal{V}_t - \mathcal{V}_0 = \int_0^t b(\mathcal{V}_t)dt + \int_0^t a(\mathcal{V}_t)d\widetilde{B}_t$, we get

$$\mathbb{E}\left(\sup_{t\in(0,\Delta)} |\mathcal{V}_t - \mathcal{V}_0|\right) \leq \Delta\mathbb{E}(|b(\mathcal{V}_0)|) + \mathbb{E}\left(\sup_{t\in(0,\Delta)}\left|\int_0^t a(\mathcal{V}_t)d\overline{B}_t\right|\right) \leq c(1 + \mathbb{E}(|\boldsymbol{Z}_0|^2_{\mathbb{R}^2})^{1/2})\Delta^{1/2}$$

using the Jensen inequality and the Burkholder-Davis-Gundy inequality.
For (i) we first see that for $x, y \in \mathbb{R}_+$ with $x < y$ holds $|\log(y) - \log(x)| = \log(y/x) = \log(1 + (y-x)/x) \leq (y-x)/x = |y-x|/x$, which implies that $|\log(y) - \log(x)| \leq |y-x|/(x \wedge y)$. We deduce that

$$\begin{aligned}\mathbb{E}(|\log(\overline{V}_1) - \log(V_0)|)^2 &\leq \mathbb{E}(|\overline{V}_1 - V_0|^2) \cdot \mathbb{E}((\overline{V}_1 \wedge V_0)^{-2}) \\ &\leq \mathbb{E}(\sup_{t\in(0,\Delta)} |V_t - V_0|^2) \cdot \mathbb{E}((\overline{V}_1 \wedge V_0)^{-2}),\end{aligned}$$

since $\overline{V}_1 - V_0 = \Delta^{-1} \int_0^\Delta V_t - V_0 dt \leq \sup_{t\in(0,\Delta)} |V_t - V_0|$. Now as $\overline{V}_1 \geq \inf_{t\in(0,\Delta)} V_t$, we get

$$\mathbb{E}((V_0 \wedge \overline{V}_1)^{-2})) \leq \mathbb{E}(\sup_{t\in(0,\Delta)} (V_t)^{-2}).$$

Analogously to (i) we can show that $\mathbb{E}(\sup_{t\in(0,\Delta)} |V_t - V_0|^2) \leq c\Delta(1 + \mathbb{E}(V_0^2))$. □

CHAPTER 4

Global survival function estimation under multiplicative measurement errors

Contents

In this chapter, we consider the estimation of the survival function $S : \mathbb{R}_+ \to [0,1]$ of a positive random variable X, defined as

$$S(x) := \mathbb{P}(X > x), \quad x \in \mathbb{R}_+,$$

under multiplicative measurement error, that is given observations $(Y_j)_{j\in[\![n]\!]}$ drawn from $Y = XU$ where the positive error random variable U is stochastically independent from X. Here, we distinguish between two cases. The case of independent observations, that is $(X_j)_{j\in[\![n]\!]}$ are identically, independent distributed and the case of stationary observations, that is $(X_j)_{j\in[\![n]\!]}$ are drawn from a

stationary discrete-time process.
Starting with the independent case, we propose in section 4.1 a survival function estimator based on the estimation of the Mellin transform of X and a regularized inverse of the Mellin transform and study the estimation strategy in terms of minimax properties and a data-driven choice of the upcoming regularization parameter.
In section, 4.2 we drop the independency assumption on the sample $(X_j)_{j\in[\![n]\!]}$ and consider samples drawn from stationary processes and study the behavior of the survival function estimator proposed in section 4.1.

4.1 Under multiplicative measurement errors

4.1.1 Introduction

In this work we are interested in estimating the unknown survival function $S : \mathbb{R}_+ \to [0, 1]$ of a positive random variable X given identically distributed copies of $Y = XU$ where X and U are independent of each other and U has a known density $g : \mathbb{R}_+ \to \mathbb{R}_+$. If X possesses a Lesbgue density $f : \mathbb{R}_+ \to \mathbb{R}_+$, unknown, then in this setting the density $f_Y : \mathbb{R}_+ \to \mathbb{R}_+$ of Y is given by

$$f_Y(y) = [f * g](y) := \int_{\mathbb{R}_+} f(x)g(y/x)x^{-1}d\lambda(x) \quad \forall y \in \mathbb{R}_+.$$

Here "$*$" denotes the multiplicative convolution. The estimation of S using a sample $(Y_j)_{j\in[\![n]\!]}$ from f_Y is thus an inverse problem called multiplicative deconvolution.

Related works A summary of related works to the estimation under multiplicative measurement errors is given in Section 3.2.1.

Organisation In Subsection 4.1.2 we propose the survival function estimator based on a spectral cut-off approach. More precisely, the estimation strategy consists of an estimator for the Mellin transform of the survival function and a regularization of the inverse Mellin transform by a spectral cut-off approach. The minimax optimality of the presented estimator is developed in Subsection 4.1.3 while a data-driven choice of the upcoming smoothing parameter, based on a penalized contrast approach, is proposed and studied in Subsection 4.1.4. The performance of the presented estimator is visualized using a Monte-Carlo simulation in Subsection 4.1.5. The proof of Subsections 4.1.3 and 4.1.4 are gathered in Subsection 4.1.7.

4.1.2 Estimation strategy

In the upcoming theory, we need to ensure that the survival function S of the sample X is square-integrable, that is $S \in \mathbb{L}^2(\mathbb{R}_+) := \mathbb{L}^2(\mathbb{R}_+, x^0)$. In this chapter, we will use the abbreviation $\|h_1\|^2 := \|h_1\|^2_{x^0}$ and $\langle h_1, h_2\rangle := \langle h_1, h_2\rangle_{x^0}$ for $h_1, h_2 \in \mathbb{L}^2(\mathbb{R}_+)$. Furthermore, to define the estimator, we additionally need the square integrability of the empirical survival function $\widehat{S}_X$ which is defined by

$$\widehat{S}_X(x) := n^{-1}\sum_{j=1}^{n} \mathbb{1}_{(0,X_j)}(x) \tag{4.1}$$

for any $x \in \mathbb{R}_+$. The following proposition shows that we can derive the square-integrability condition for both functions by a moment condition.

Proposition 4.1.1:
Let $\mathbb{E}_f(X^{1/2}) < \infty$. Then, $S \in \mathbb{L}^1(\mathbb{R}_+, x^{-1/2}) \cap \mathbb{L}^2(\mathbb{R}_+)$. If additionally $\mathbb{E}_f(X_1) < \infty$ then $\widehat{S}_X \in \mathbb{L}^1(\mathbb{R}_+, x^{-1/2}) \cap \mathbb{L}^2(\mathbb{R}_+)$ almost surely.

The proof of Proposition 4.1.1 can be found in the Subsection 4.1.7. Following the steps of section 3.1 and assuming now that $\mathbb{E}_f(X_1^{1/2}) < \infty$ we have for $k \in \mathbb{R}_+$ that $\mathcal{M}_{1/2}[S]\mathbb{1}_{[-k,k]} \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ and thus

$$S_k(x) := \mathcal{M}^{\dagger}_{1/2}[\mathcal{M}_{1/2}[S]\mathbb{1}_{[-k,k]}](x) = \frac{1}{2\pi}\int_{-k}^{k} x^{-1/2-it}\mathcal{M}_{1/2}[S](t)d\lambda(t), \quad x \in \mathbb{R}_+, \tag{4.2}$$

is an approximation of S in the $\mathbb{L}^2(\mathbb{R}_+)$ sense, that is, $\|S_k - S\|^2 \to 0$ for $k \to \infty$. Now, applying the property of the Mellin transform for survival functions, compare Proposition 2.2.4, we obtain that $\mathcal{M}_{1/2}[S](t) = (1/2+it)^{-1}\mathcal{M}_{3/2}[f](t)$ for all $t \in \mathbb{R}$.
If we in addition to $\mathbb{E}_f(X^{1/2}) < \infty$ assume that $\mathbb{E}_g(U^{1/2}) < \infty$, then $\mathbb{E}_{f_Y}(Y^{1/2}) < \infty$, and $f_Y, g \in \mathbb{L}^1(\mathbb{R}, x^{1/2})$. The convolution theorem implies then $\mathcal{M}_{3/2}[f_Y] = \mathcal{M}_{3/2}[f]\mathcal{M}_{3/2}[g]$. Under the mild assumption that for all $t \in \mathbb{R}, \mathcal{M}_{3/2}[g](t) \neq 0$ we can rewrite equation (4.2) in the following form

$$S_k(x) = \frac{1}{2\pi}\int_{-k}^{k} x^{-1/2-it}\frac{\mathcal{M}_{3/2}[f_Y](t)}{(1/2+it)\mathcal{M}_{3/2}[g](t)}d\lambda(t). \tag{4.3}$$

To derive an estimator from (4.3) we use the empirical Mellin transform given by $\widehat{\mathcal{M}}(t) := \widehat{\mathcal{M}}_{3/2}(t) = n^{-1}\sum_{j=1}^{n} Y_j^{1/2+it}$ as an unbiased estimator of $\mathcal{M}_{3/2}[f_Y](t)$ for all $t \in \mathbb{R}$. Keeping in mind that $|\widehat{\mathcal{M}}(t)| \leq |\widehat{\mathcal{M}}(0)| < \infty$, it is sufficient to assume that $\int_{-k}^{k} |(1/2+it)\mathcal{M}_{3/2}[g](t)|^{-2}dt < \infty$ for all $k \in \mathbb{R}_+$ to ensure that the spectral cut-off estimator

$$\widehat{S}_k(x) := \frac{1}{2\pi}\int_{-k}^{k} x^{-1/2-it}\frac{\widehat{\mathcal{M}}(t)}{(1/2+it)\mathcal{M}_{3/2}[g](t)}d\lambda(t), \quad k, x \in \mathbb{R}_+, \tag{4.4}$$

is well-defined. Up to now, we impose two minor conditions on the Mellin transform of the error density g which we want to collect in the following assumption,

$$\forall\, t \in \mathbb{R} : \mathcal{M}_{3/2}[g](t) \neq 0 \quad \text{and} \quad \forall\, k \in \mathbb{R}_+ : \int_{-k}^{k} |(1/2+it)\mathcal{M}_{3/2}[g](t)|^{-2}d\lambda(t) < \infty. \tag{[G0]}$$

Remark 4.1.2 (The Mellin transform of the empirical survival function)**:**
It is worth stressing out, that for each $t \in \mathbb{R}$ the estimator

$$(1/2 + it)^{-1}\widehat{\mathcal{M}}_X(t) := (1/2 + it)^{-1} n^{-1} \sum_{j=1}^{n} X_j^{1/2+it}$$

is an unbiased estimator of $\mathcal{M}_{1/2}[S](t)$. Furthermore, there is a special link between the empirical survival function and the estimator $(1/2 + it)^{-1/2}\widehat{\mathcal{M}}(t)$. In fact, Proposition 4.1.1 ensures that $\widehat{S}_X \in \mathbb{L}^1(\mathbb{R}_0, x^{-1/2})$. Consequently, the Mellin transform of $\widehat{S}_X$ is well-defined and

$$\mathcal{M}_{1/2}[\widehat{S}_X](t) = (1/2 + it)^{-1}\widehat{\mathcal{M}}_X(t) \tag{4.5}$$

for all $t \in \mathbb{R}$.

In analogy to Proposition 3.2.1 we can prove the following risk bound whose proof can be found in Section 4.1.7.

Proposition 4.1.3 (Upper bound of the risk)**:**
Assume that $\mathbb{E}_{f_Y}(Y) < \infty$ and that **[G0]** holds. Then for any $k \in \mathbb{R}_+$,

$$\mathbb{E}^n_{f_Y}(\|\widehat{S}_k - S\|^2) \leq \|S - S_k\|^2 + \mathbb{E}_{f_Y}(Y_1)\frac{\Delta_g(k)}{n}$$

where $\Delta_g(k) = (2\pi)^{-1}\int_{(-k,k)} |(1/2 + it)\mathcal{M}_{3/2}[g](t)|^{-2} d\lambda(t)$.

By dominated convergence the squared bias term $\|S - S_k\|^2$ is monotonically decreasing in $k \in \mathbb{R}_+$ while $\Delta_g(k)n^{-1}$ is monotonically increasing. Therefore, any sequence $(k_n)_{n\in\mathbb{N}}$ with $k_n \to \infty$ and $\Delta_g(k_n)n^{-1} \to 0$ leads to a consistent survival function estimator $\widehat{S}_{k_n}$ which is captured in the following Corollary.

Corollary 4.1.4 (Consistency)**:**
If $(k_n)_{n\in\mathbb{N}}$ is chosen such that $k_n \to \infty$ for $n \to \infty$ and $\Delta_g(k_n)n^{-1} \to 0$, the consistency of $\widehat{S}_{k_n}$, that is,

$$\mathbb{E}^n_{f_Y}(\|\widehat{S}_{k_n} - S\|^2) \to 0, \qquad n \to \infty$$

is implied.

For a more sophisticated analysis of the variance's growth, we need to consider more specific assumptions on the error density g. More precisely, the growth of Δ_g is determined by the decay of the Mellin transform of g.

4.1.3 Minimax theory

As before, we characterize the error density g by the decay of the Mellin transform $\mathcal{M}_c[g]$ to determine the exact growth of the variance term. That is, we mainly focus on the case of SMOOTH ERROR DENSITIES, introduced in Definition 3.2.4. As a reminder, an error density is called smooth if there exists $c, C, \gamma \in \mathbb{R}_+$ such that

$$c(1 + t^2)^{-\gamma/2} \leq |\mathcal{M}_c[g](t)| \leq C(1 + t^2)^{-\gamma/2} \text{ for } t \in \mathbb{R}. \tag{[G1]}$$

Parametric rate In the special case, that g satisfies **[G1]** with $\gamma \in (0, 1/2)$, there is a choice of the parameter $k \in \mathbb{R}_+$ which leads to a parametric rate up to a log-term. It is worth pointing out, that this choice can be done independently of the precise decay of the bias term $\|S - S_k\|^2$. This can be accomplished by the elementary bound

$$\|S - S_k\|^2 = \frac{1}{\pi} \int_k^\infty |\mathcal{M}_{1/2}[S](t)|^2 d\lambda(t) \leq k^{-1} \mathbb{E}_f(X_1^{1/2})^2$$

where we exploit $\mathcal{M}_{1/2}[S](t) = (1/2 + it)^{-1} \mathcal{M}_{3/2}[f](t)$ and the bound $|\mathcal{M}_{3/2}[f](t)| \leq \mathbb{E}_f(X_1^{1/2})$. The different cases are collected in the following Proposition whose proof is omitted.

Proposition 4.1.5 (Parametric risk bound)**:**
Let $\mathbb{E}(Y) < \infty$ and let g satisfy **[G1]** with $\gamma \in (0, 1/2)$. Then, there exists $C(g) \in \mathbb{R}_+$ such that for all $k \in \mathbb{R}_+^+$

$$\Delta_g(k) \leq \begin{cases} C(g) & , \gamma < 1/2; \\ C(g) \log(k) & , \gamma = 1/2. \end{cases}$$

Choosing now $k = n$, then for $\gamma = 1/2$, we end up with parametric rate up to a log-term, that is

$$\mathbb{E}^n_{f_Y}(\|\widehat{S}_n - S\|^2) \leq C(\mathbb{E}_f(X_1), g) \frac{\log(n)}{n}$$

while if $\gamma < 1/2$ we accomplish a parametric rate, that is

$$\mathbb{E}^n_{f_Y}(\|\widehat{S}_n - S\|^2) \leq \frac{C(\mathbb{E}_f(X_1), g)}{n}$$

where $C(\mathbb{E}_f(X_1), g)$ depends of $\mathbb{E}_f(X_1), \mathbb{E}_g(U_1)$ and the constants in **[G1]**.

As demonstrated by Proposition 4.1.5, for $\gamma \leq 1/2$ choosing $k = n$, independent of $f \in \mathbb{L}^2(\mathbb{R}_+)$, we can derive a parametric rate, respectively almost parametric rate. For the case that $\gamma > 1/2$ we get under assumption **[G1]** for $c = 3/2$ that $c_g \leq k^{2\gamma - 1} \leq \Delta_g(k) \leq C_g k^{2\gamma-1}$ for positive constants $c_g, C_g > 0$ possibly dependent on g. Thus the variance term is monotonically increasing and unbounded leading us to the usual bias-variance conflict.

Non-parametric rates To control the bias-term we make use again of regularity spaces which characterize the decay of the Mellin transform similarly to Sobolev spaces for deconvolution problems. For $s \in \mathbb{R}_+$ consider the MELLIN-SOBOLEV SPACE, defined in Definition 3.1.3 by

$$\mathbb{W}^s_{1/2}(\mathbb{R}_+) := \{h \in \mathbb{L}^2(\mathbb{R}_+) : |h|_s^2 := \|(1 + t^2)^{s/2} \mathcal{M}_{1/2}[h]\|^2_{\mathbb{R}} < \infty\} \tag{4.6}$$

and the corresponding ellipsoids with $L \in \mathbb{R}_+$ by $\mathbb{W}^s_{1/2}(L) := \{h \in \mathbb{W}^s_{1/2}(\mathbb{R}_+) : |h|_s^2 \leq L\}$. For $S \in \mathbb{W}^s_{1/2}(L)$ we deduce that $\|S - S_k\|^2 \leq L k^{-2s}$. Setting

$$\mathbb{S}^s_{1/2}(L) := \{S \in \mathbb{W}^s_{1/2}(L) : \text{S survival function}, \mathbb{E}_f(X_1) \leq L\}$$

and the previous discussion leads to the following statement whose proof is omitted.

Proposition 4.1.6 (Upper bound for the minimax risk)**:**
Let $\mathbb{E}_g(U_1) < \infty$. Then under the assumption **[G1]** with $\gamma > 1/2$,

$$\sup_{S \in \mathbb{S}^s_{1/2}(L)} \mathbb{E}^n_{f_Y}(\|S - \widehat{S}_{k_n}\|)^2 \leq C(L, g, s) n^{-2s/(2s+2\gamma-1)}$$

for the choice $k_n := n^{1/(2s+2\gamma-1)}$.

We will now show, that the rates presented in Proposition 4.1.6 are optimal in the sense, that there exists no estimator based on the i.i.d. sample $Y_1, \dots, Y_n$ that can reach uniformly over $\mathbb{S}^s_{1/2}(L)$ a better rate. This implies that the estimator $\widehat{S}_{k_n}$ presented in 4.1.6 is minimax-optimal.
For technical reasons we need an additional assumption on the error density g. Let us assume that there are constants $c_g, C_g, \gamma \in \mathbb{R}_+$ such that

$$\forall x > 1 : g(x) = 0 \text{ and } c_g(1+t^2)^{-\gamma/2} \leq |\mathcal{M}_{1/2}[g](t)| \leq C_g(1+t^2)^{-\gamma/2} \text{ for } t \in \mathbb{R}. \qquad \textbf{([G1'])}$$

With this additional assumption we can show the following theorem whose proof can be found in Subsection 4.1.7.

Theorem 4.1.7 (Lower bound for the minimax risk)**:**
Let $s, \gamma \in \mathbb{N}$ and assume that **[G1]** and **[G1']** hold. Then there exist constants $c_g, L_{s,g} > 0$ such that for all $L \geq L_{s,g}$, $n \in \mathbb{N}$, and for any estimator $\widehat{S}$ of S based on an i.i.d. sample $Y_1, \dots, Y_n$,

$$\sup_{S \in \mathbb{S}^s_{1/2}(L)} \mathbb{E}^n_{f_Y}(\|\widehat{S} - S\|^2) \geq c_g n^{-2s/(2s+2\gamma-1)}.$$

For the multiplicative censoring model, that is, U is uniformly-distributed on $(0, 1)$, the assumptions **[G1]** and **[G1']** hold true. Furthermore, despite the fact that the choice of k_n in Proposition 4.1.6 is not dependent on the explicit survival function $S \in \mathbb{S}^s_{1/2}(L)$, it is still dependent on the regularity parameter $s \in \mathbb{R}_+$ of the unknown survival function S, which is unknown, too. While it is tempting to set the regularity parameter $s \in \mathbb{R}_+$ to a fixed value and interpret this as an additional model assumption, the discussion below motivates that this might deliver worse rates.

Faster rates Revisiting Section 3.1.3 and using the survival function formula for the Mellin transform, we get that the families of Gamma and Log-Normal distributions satisfy

$$\|S - S_k\|^2 = \frac{1}{\pi} \int_k^\infty \frac{|\mathcal{M}_{3/2}[f](t)|^2}{1/4 + t^2} d\lambda(t) \leq C_f k^{-2} \exp(-\alpha k^a)$$

for $\alpha \in \mathbb{R}_+$, dependent on f and $a = 1$ for the Gamma distribution and $a = 2$ for the Log-Normal distribution. Now choosing $k_n = \log(n)^{1/a}/\alpha^{-1}$ we can ensure that for both cases

$$\mathbb{E}^n_{f_Y}(\|\widehat{S}_{k_n} - S\|^2) \leq C_f \frac{\log(n)^{(2\gamma-1)/a}}{n}$$

leading to a faster rate than provided by Proposition 4.1.6 for any choice of $s \in \mathbb{R}_+$. In the next section, we therefore present a data-driven method in order to choose the parameter $k \in \mathbb{R}_+$ based on the sample $Y_1, \dots, Y_n$.

4.1.4 Data-driven method

We now present a data-driven method based on a penalized contrast approach. In fact, we consider the case where **[G1]** holds with $\gamma > 1/2$, since in the case $\gamma \leq 1/2$ we already presented a choice of the parameter $k \in \mathbb{R}_+$, independent on the density f, which achieves an almost parametric rate. In the case $\gamma > 1/2$ the variance term is increasing, unbounded and characterized by the growth of Δ_g. Our aim is now to define an estimator $\widehat{k}_n$ which mimics the behavior of

$$k_n := \arg\min_{k \in \mathcal{K}_n} \|S - S_k\|^2 + \frac{\mathbb{E}_{f_Y}(Y_1) C_g}{2\pi n} k^{2\gamma - 1}$$

for a suitable large set of parameters $\mathcal{K}_n \subset \mathbb{R}_+$. Considering the result of Proposition 4.1.6, and the fact that $\|S - S_k\|^2 \leq k^{-1}\mathbb{E}_f(X_1^{1/2})^2$, which we have seen in the paragraph about the parametric case, we can ensure that the set $\mathcal{K}_n := \{k \in \{1, \dots, n\} : \Delta_g(k) \leq n\}$ is suitably large enough. Starting with the bias term we see that $\|S - S_k\|^2 = \|S\|^2 - \|S_k\|^2$ behaves like $-\|S_k\|^2$. Furthermore, for $k \in \mathcal{K}_n$ we define the penalty term $\mathrm{pen}(k) = \chi \sigma_Y \Delta_g(k) n^{-1}$, $\sigma_Y := \mathbb{E}_{f_Y}(Y_1)$, which shall mimic the behavior of the variance term. Exchanging $-\|S_k\|^2$ and $\mathbb{E}_{f_Y}(Y_1)$ with their empirical counterparts $-\|\widehat{S}_k\|^2$ and $\widehat{\sigma}_Y := n^{-1}\sum_{j=1}^n Y_j$ we define a fully data-driven model selection $\widehat{k}$ by

$$\widehat{k} := \arg\min_{k \in \mathcal{K}_n} -\|\widehat{S}_k\|^2 + \widehat{\mathrm{pen}}(k) \quad \text{where} \quad \widehat{\mathrm{pen}}(k) := 2\chi \widehat{\sigma}_Y \Delta_g(k) n^{-1} \tag{4.7}$$

for $\chi > 0$. The following theorem shows that this procedure is adaptive up to a negligible term.

Theorem 4.1.8 (Data-driven choice of $k \in \mathbb{R}_+$)**:**
Let g satisfy **[G1]** with $\gamma > 1/2$ and $\|xg\|_\infty < \infty$. Assume further that $\mathbb{E}_{f_Y}(Y_1^{5/2}) < \infty$. Then for $\chi > 48$,

$$\mathbb{E}^n_{f_Y}(\|S - \widehat{S}_{\widehat{k}}\|^2) \leq 6 \inf_{k \in \mathcal{K}_n} \left(\|S - S_k\|^2 + \mathrm{pen}(k)\right) + \frac{C(g, f)}{n}$$

where $C(g, f) > 0$ is a constant depending on χ, the error density g, $\mathbb{E}_{f_Y}(Y_1^{5/2})$, σ_Y.

The proof of Theorem 4.1.8 is postponed to Section 4.1.7. The assumption that $\|xg\|_\infty < \infty$ is rather weak because it is satisfied for the common densities considered. Under the assumptions of Theorem 4.1.8 and for the case $S \in \mathbb{S}^s_{1/2}(L)$, we can ensure that $s > 1/2$. Then we have $k_n := \lfloor n^{1/(2s+2\gamma-1)} \rfloor \in \mathcal{K}_n$ and thus

$$\mathbb{E}^n_{f_Y}(\|S - \widehat{S}_{\widehat{k}}\|^2) \leq C(f, g) n^{-2s/(2s+2\gamma-1)}.$$

4.1.5 Numerical results

Let us illustrate the performance of the fully-data driven estimator $\widehat{S}_{\widehat{k}}$ defined in 4.4 and 4.7. To do so, we consider the following densities whose corresponding survival function will be estimated.

(*i*) Gamma distribution: $f_1(x) = \frac{0.5^4}{\Gamma(4)} x^3 \exp(-0.5x) \mathbb{1}_{\mathbb{R}_+}(x)$,

(*ii*) Weibull distribution: $f_2(x) = 2x \exp(-x^2) \mathbb{1}_{\mathbb{R}_+}(x)$,

(*iii*) Beta distribution: $f_3(x) = \frac{1}{560}(0.5x)^3(1 - 0.5x)^4 \mathbb{1}_{(0,2)}(x)$ and (*iv*) Log-Gamma distribution: $f_4(x) = \frac{x^{-4}}{6}\log(x)^3\mathbb{1}_{(1,\infty)}(x)$.

For the distribution of the error density, we consider the following three cases

a) Uniform distribution: $g_1(x) = \mathbb{1}_{[0,1]}(x)$,

b) Symmetric noise: $g_2(x) = \mathbb{1}_{(1/2,3/2)}(x)$ and

c) Beta distribution: $g_3(x) = 2(1-x)\mathbb{1}_{(0,1)}(x)$.

We see that g_1 and g_2 fulfill **[G1]** with the parameter $\gamma = 1$ and g_3 fulfills it with $\gamma = 2$. Due to the fact that for the true survival function holds $S(x) \in [0,1], x \in \mathbb{R}$, we can improve the estimator $\widehat{S}_{\widehat{k}}$ by defining

$$\widetilde{S}_{\widehat{k}}(x) := \begin{cases} 0 & , \widehat{S}_{\widehat{k}}(x) \leq 0; \\ \widehat{S}_{\widehat{k}}(x) & , \widehat{S}_{\widehat{k}}(x) \in [0,1]; \\ 1 & , \widehat{S}_{\widehat{k}}(x) \geq 1. \end{cases}$$

The resulting estimator $\widetilde{S}_{\widehat{k}}$ has a smaller risk then $\widehat{S}_{\widehat{k}}$, since $\|\widetilde{S}_{\widehat{k}} - S\|^2 \leq \|\widehat{S}_{\widehat{k}} - S\|^2$. On the other hand, the estimator has the desired property that it is $[0,1]$-valued. Nevertheless, it is difficult ensure that $\widetilde{S}_{\widehat{k}}(0) = 1$ and that $\widetilde{S}_{\widehat{k}}$ is monotone decreasing. Although there are many procedures to guarantee the monotonicity of an estimator $\widehat{S}$ and the property $\widehat{S}(0) = 1$, the presented theoretical results of this work are not applicable to the modified estimators.
We will now use a Monte-Carlo simulation to visualize the properties of the estimator $\widetilde{S}_{\widehat{k}}$ and discuss whether the numerical simulated behavior of the estimator coincides with the theoretical predictions. Afterwards, we will construct a survival estimator $\widehat{S}$ based on $\widehat{S}_{\widehat{k}}$ which is in fact a survival function, keeping in mind that the theoretical results of this work do not apply for this estimator.

	$n = 500$	$n = 1000$	$n = 2000$
(i)	1.31	0.75	0.23
(ii)	0.19	0.09	0.04
(iii)	0.21	0.12	0.06
(iv)	1.10	0.34	0.20

Figure 4.1: The entries showcase the MISE (scaled by a factor of 100) obtained by Monte-Carlo simulations each with 200 iterations. We take a look at different densities and varying sample size. The error density is chosen as a) in each case.

In Figure 4.1 we can see that for an increasing sample size, the variance of the estimator seems to decrease. Also, increasing the sample size allows for the estimator to choose bigger values for $\widehat{k}$, which on the other hand decreases the bias induced by the approximation step. Next let us consider different error densities. In Figure 4.2 we can see that reconstruction of the survival function with error density a) and b) seem to be of same complexity while the reconstruction with error density c) seems to be more difficult. This behavior is predicted by the theoretical results because for **[G1]**, a) and b) share the same parameter $\gamma = 1$ while c) has the parameter $\gamma = 2$.

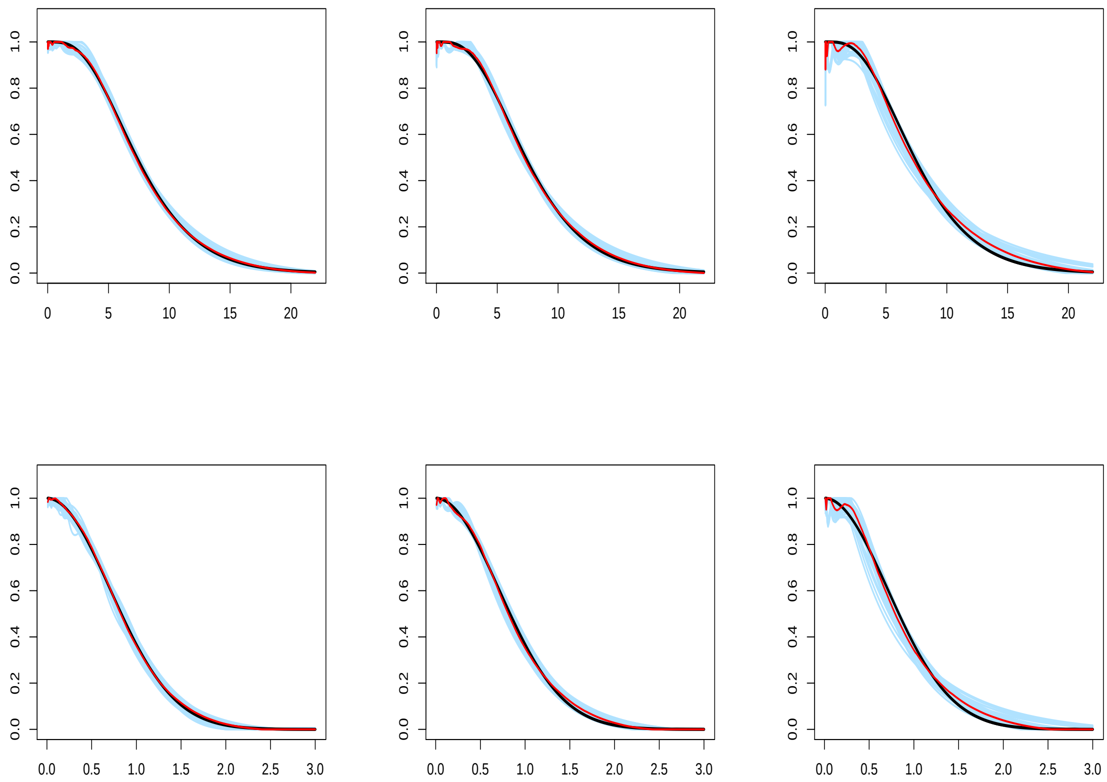

Figure 4.2: Considering the estimators $\widetilde{S}_{\widehat{k}}$, we depict 50 Monte-Carlo simulations with varying error density a) (left), b) (middle) and c) (right) for (i) (top), (ii) (bottom) with $n = 1000$. The true survival function S is given by the black curve while the red curve is the point-wise empirical median of the 50 estimates.

Heuristic estimator We will now modify the estimator $\widehat{S}_{\widehat{k}}$ such that the resulting estimator $\widehat{S}$ is a survival function. To do so, we see that for any $k \in \mathbb{R}$ and $x \in \mathbb{R}_+$,

$$\widehat{S}_k(x) = \frac{1}{2\pi}\int_{-k}^{k} x^{-1/2-it}\frac{\widehat{\mathcal{M}}(t)}{\mathcal{M}_{3/2}[g](t)}(1/2+it)^{-1}d\lambda(t) = (\widehat{p}_k * g_u)(x),$$

where $g_u(x) = \mathbb{1}_{(0,1)}(x)$ the density of the uniform distribution on the interval $(0,1)$ and $\widehat{p}_k(x) = (2\pi)^{-1}\int_{-k}^{k} x^{-1/2-it}\frac{\widehat{\mathcal{M}}(t)}{\mathcal{M}_{3/2}[g](t)}dt$. Exploiting the definition of the multiplicative convolution we see that for any $x \in \mathbb{R}_+$,

$$\widehat{S}_k(x) = \int_x^{\infty} \widehat{p}_k(y)y^{-1}d\lambda(y).$$

This motivates the following construction of the survival function estimator. First, exchanging $\widehat{p}_k$ with $(\widehat{p}_k(x))_+$ ensures the monotonicity and the positivity of our estimator. The final estimator is then defined as

$$\widehat{S}(x) := \widetilde{S}(x)/\widetilde{S}(0+), \quad \text{where } \widetilde{S}(x) := \int_x^\infty (\widehat{p}_{\widehat{k}}(y))_+ y^{-1} d\lambda(y), \text{ for any } y \in \mathbb{R}_+,$$

where $0+$ denotes a positive real number very close to 0. Since our estimator is not defined in 0 we cannot normalize it with $\widetilde{S}(0)$.

Let us now illustrate the behavior of the heuristic estimator $\widehat{S}$ for an increasing number of observations compared to the estimator $\widetilde{S}_{\widehat{k}}$.

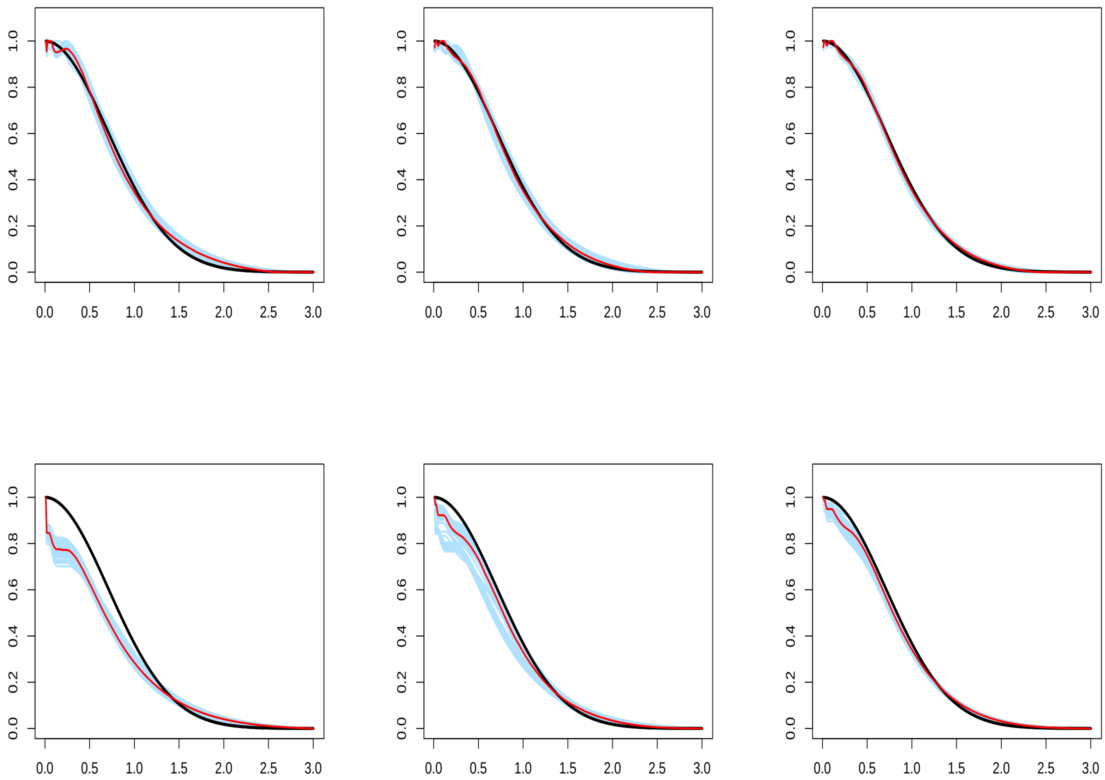

Figure 4.3: Considering the estimators $\widetilde{S}_{\widehat{k}}$ (top) and $\widehat{S}$ (bottom), we depict 50 Monte-Carlo simulations with varying sample size $n = 500$ (left), $n = 1000$ (middle) and $n = 2000$ (right) in the case (ii) with error density (b). The true survival function S is given by the black curve while the red curve is the point-wise empirical median of the 50 estimates.

Although the estimator $\widehat{S}$ is a survival function, the modification of the estimator seems to introduce an additional bias to the estimation. Based on the numerical study, this additional bias seems to decrease for n large enough. Nevertheless, we want to stress out that this modification is purely heuristic.

4.1.6 Conclusion

In this section, we have constructed a family of estimators $(\widehat{S}_k)_{k \in \mathbb{R}_+}$ for the survival function $S : \mathbb{R}_+ \to [0, 1]$ of a positive random variable X based on i.i.d. observations $(Y_j)_{j \in [\![n]\!]}$ with multiplicative measurement error.

We studied our estimator in terms of the mean integrated squared error, and showed that there exists a choice of the cut-off parameter $k_o \in \mathbb{R}_+$ which solves the squared bias-variance conflict uniformly over Sobolev-Mellin ellipsoids. The resulting rate of convergence has been shown to be minimax-optimal by stating a lower bound result for the estimation of the survival function S based on the observations $(Y_j)_{j\in[\![n]\!]}$ over the Mellin-Sobolev ellipsoids.
Although the choice k_o leads to a minimax optimal estimator $\widehat{S}_{k_o}$, the explicit choice k_o is still dependent on unknown smoothness parameters of the Mellin-Sobolev ellipsoid and is thus not feasible for practical applications. We therefore proposed a fully-data driven choice of the cut-off parameter based on a model selection approach and proved its adaptivity.
A Monte-Carlo simulation then visualized the reasonable behavior of our estimation strategy. Furthermore, we presented an heuristic approach to ensure that our estimator is indeed a survival function. Theoretical properties of this heuristic approach are left for further research.

4.1.7 Proofs

Proof of Proposition 4.1.1. Let us assume that $\mathbb{E}_f(X^{1/2}) < \infty$. Then, $f \in \mathbb{L}^1(\mathbb{R}_+, x^{1/2})$ and thus $S \in \mathbb{L}^1(\mathbb{R}_+, x^{-1/2})$ because

$$\int_{\mathbb{R}_+} x^{-1/2} S(x) d\lambda(x) = \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} x^{-1/2}\mathbb{1}_{(0,x)}(y) f(y) d\lambda(y) d\lambda(x) = \int_{\mathbb{R}_+} 2y^{1/2} f(y) d\lambda(y) < \infty.$$

By the generalized Minkowski inequality, cf. Tsybakov (2009), we have

$$\begin{aligned}\int_{\mathbb{R}_+} S^2(x) d\lambda(x) &= \int_{\mathbb{R}_+} \left(\int_{\mathbb{R}_+} \mathbb{1}_{(0,x)}(y) f(x) d\lambda(x)\right)^2 d\lambda(y)\\ &\le \left(\int_{\mathbb{R}_+}\left(\int_{\mathbb{R}_+} \mathbb{1}_{(0,x)}(y) f^2(x) d\lambda(y)\right)^{1/2} d\lambda(x)\right)^2 = \left(\int_{\mathbb{R}_+} x^{1/2} f(x) d\lambda(x)\right)^2,\end{aligned}$$

which implies that $S \in \mathbb{L}^2(\mathbb{R}_+, x^0)$. For the second part we see that $\mathbb{E}^n_f(\|\widehat{S}_X\|_{\mathbb{L}^1(\mathbb{R}_+,x^{-1/2})}) = \int_0^\infty x^{-1/2} S_X(x) dx < \infty$. Since $\mathbb{E}_f(X_1) < \infty$, we get

$$\mathbb{E}^n_f(\|\widehat{S}_X\|^2_{x^0}) \le \int_{\mathbb{R}_+} \mathbb{E}(\mathbb{1}_{(x,\infty)}(X_1)) d\lambda(x) = \int_{\mathbb{R}_+} S(x) d\lambda(x) = \mathbb{E}_f(X_1) < \infty.$$

This implies that $\widehat{S}_X \in \mathbb{L}^1(\mathbb{R}_+, x^{-1/2}) \cap \mathbb{L}^2(\mathbb{R}_+, x^0)$ almost surely. □

Proof of Proposition 4.1.3. Let $k \in \mathbb{R}_+$. Since $\mathcal{M}_{1/2}[S - S_k](t) = 0$ for $|t| \le k$ we get by application of the Plancherel identity, $\langle S - S_k, S_k - \widehat{S}_k\rangle = \frac{1}{2\pi}\int_{-k}^k \mathcal{M}_{1/2}[S - S_k](t)\mathcal{M}_{1/2}[S_k - \widehat{S}_k](-t) d\lambda(t) = 0$ and thus $\|S - \widehat{S}_k\|^2 = \|S - S_k\|^2 + \|S_k - \widehat{S}_k\|^2$. Again by application of the Plancherel identity, cf. Proposition 2.3.5, and the Fubini-Tonelli theorem,

$$\mathbb{E}^n_{f_Y}(\|\widehat{S}_k - S_k\|^2) = \frac{1}{2\pi}\int_{-k}^k \frac{\mathbb{V}\mathrm{ar}(\widehat{\mathcal{M}}(t))}{|(1/2 + it)\mathcal{M}_{3/2}[g](t)|^2} d\lambda(t).$$

Now the theorem follows since $\mathbb{V}\mathrm{ar}^n_{f_Y}(\widehat{\mathcal{M}}(t)) \le n^{-1}\mathbb{E}_{f_Y}(Y_1)$. □

Proof of Theorem 4.1.7. We first outline the main steps of the proof. We will construct a family of functions in $\mathbb{S}^s_{1/2}(L)$ by a perturbation of the survival function $S_o : \mathbb{R}_+ \to \mathbb{R}_+$ with small "bump", such that their $\mathbb{L}^2$-distance and the Kullback-Leibler divergence of their induced distributions can be bounded from below and above, respectively. The claim follows by applying Theorem 2.5 in Tsybakov (2009). We use the following construction.
Let $\psi \in C_0^\infty(\mathbb{R})$ be a function with support in $[0,1]$ and $\int_0^1 \psi(x)d\lambda(x) = 0$. For each $K \in \mathbb{N}$, to be selected below, and $k \in [\![0,K]\!]$ we define the bump-functions $\psi_{k,K}(x) := \psi(xK - K - k)$, $x \in \mathbb{R}$. For $j \in \mathbb{N}_0$ we set the finite constant $C_{j,\infty} := \max(\|\psi^{(l)}\|_\infty, l \in [\![0,j]\!])$. Using the operator $\mathrm{S} : C_0^\infty(\mathbb{R}) \to C_0^\infty(\mathbb{R})$ with $\mathrm{S}[f](x) = -xf^{(1)}(x)$ for all $x \in \mathbb{R}$, compare Proposition 2.3.12, we define for $j \in \mathbb{N}$ the function $\psi_{k,K,j}(x) := \mathrm{S}^j[\psi_{k,K}](x) = (-1)^j \sum_{i=1}^j c_{i,j} x^i K^i \psi^{(i)}(xK-K-k)$ for $x \in \mathbb{R}_+$ and $c_{i,j} \geq 1$. For a bump-amplitude $\delta > 0, \gamma \in \mathbb{N}$ and a vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K) \in \{0,1\}^K$ we define

$$S_{\boldsymbol{\theta}}(x) = S_o(x) + \delta K^{-s-\gamma+1} \sum_{k=0}^{K-1} \theta_{k+1}\psi_{k,K,\gamma-1}(x) \text{ where } S_o(x) := \exp(-x). \tag{4.8}$$

Taking the negative sign of the derivative of this function leads to the density

$$f_{\boldsymbol{\theta}}(x) = f_o(x) + \delta K^{-s-\gamma+1} \sum_{k=0}^{K-1} \theta_{k+1}\psi_{k,K,\gamma}(x)x^{-1} \text{ where } f_o(x) := \exp(-x). \tag{4.9}$$

Until now, we did not give a sufficient condition to ensure that our constructed functions $\{S_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \{0,1\}^K\}$ are in fact survival functions. We do this by stating conditions such that the family $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \{0,1\}^K\}$ is a family of densities.

Lemma 4.1.9:
Let $0 < \delta < \delta_o(\psi, \gamma) := \exp(-2)2^{-\gamma}(C_{\gamma,\infty}c_\gamma)^{-1}$. Then for all $\boldsymbol{\theta} \in \{0,1\}^K$, $f_{\boldsymbol{\theta}}$ is a density.

Furthermore, it is possible to show that these survival functions all lie inside the ellipsoids $\mathbb{S}^s_{1/2}(L)$ for L big enough. This is captured in the following lemma.

Lemma 4.1.10:
Let $s \in \mathbb{N}$. Then, there is $L_{s,\gamma,\delta} > 0$ such that S_o and any $S_{\boldsymbol{\theta}}$ as in Equation (4.9) with $\boldsymbol{\theta} \in \{0,1\}^K$, $K \in \mathbb{N}$, belong to $\mathbb{S}^s_{1/2}(L_{s,\gamma,\delta})$.

For sake of simplicity we denote for a function $\varphi \in \mathbb{L}^2(\mathbb{R}_+) \cap \mathbb{L}^1(\mathbb{R}_+, x^{-1/2})$ the multiplicative convolution with g by $\widetilde{\varphi} := \varphi * g$. Next, we see that for $y_2 \geq y_1 > 0$,

$$\widetilde{f}_o(y_1) = \int_0^\infty g(x)x^{-1}\exp(-y_1/x)dx \geq \int_0^\infty g(x)x^{-1}\exp(-y_2/x)dx = \widetilde{f}_o(y_2)$$

and thus $\widetilde{f}_o$ is monotonically decreasing. Additionally, we have that $\widetilde{f}_o(2) > 0$, since otherwise $g = 0$ almost everywhere. Exploiting Varshamov-Gilbert's lemma (cf. Tsybakov (2009)) in Lemma 4.1.11 we show that there exists $M \in \mathbb{N}$ with $M \geq 2^{K/8}$ and a subset $\{\boldsymbol{\theta}^{(0)}, \dots, \boldsymbol{\theta}^{(M)}\}$ of $\{0,1\}^K$ with $\boldsymbol{\theta}^{(0)} = (0, \dots, 0)$ such that for all $j, l \in [\![0,M]\!]$, $j \neq l$, the $\mathbb{L}^2$-distance and the Kullback-Leibler divergence are bounded for $K \geq K_o(\gamma, \psi)$.

Lemma 4.1.11:
Let $K \geq \max\{K_o(\psi,\gamma), 8\}$. Then there exists a subset $\{\boldsymbol{\theta}^{(0)}, \dots, \boldsymbol{\theta}^{(M)}\}$ of $\{0,1\}^K$ with $\boldsymbol{\theta}^{(0)} = (0,\dots,0)$ such that $M \geq 2^{K/8}$ and for all $j, l \in [\![0, M]\!], j \neq l$,

$$(i) \quad \|S_{\boldsymbol{\theta}^{(j)}} - S_{\boldsymbol{\theta}^{(l)}}\|^2 \geq \frac{\|\psi^{(\gamma-1)}\|^2\delta^2}{16} K^{-2s} \text{ and}$$
$$(ii) \quad \mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}^{(j)}}, \widetilde{f}_{\boldsymbol{\theta}^{(0)}}) \leq \frac{C_1(g)\|\psi\|^2}{\widetilde{f}_o(2)\log(2)}\delta^2 \log(M) K^{-2s-2\gamma+1}$$

where KL is the Kullback-Leibler-divergence.

Selecting $K = \lceil n^{1/(2s+2\gamma-1)} \rceil$ delivers

$$\frac{1}{M}\sum_{j=1}^{M} \mathrm{KL}((\widetilde{f}_{\boldsymbol{\theta}^{(j)}})^{\otimes n}, (\widetilde{f}_{\boldsymbol{\theta}^{(0)}})^{\otimes n}) = \frac{n}{M}\sum_{j=1}^{M} \mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}^{(j)}}, \widetilde{f}_{\boldsymbol{\theta}^{(0)}}) \leq c_{\psi,\delta,g,\gamma,f_o}\log(M)$$

where $c_{\psi,\delta,g,\gamma,f_o} < 1/8$ for all $\delta \leq \delta_1(\psi, g, \gamma, f_o)$ and $M \geq 2$ for all $n \geq n_{s,\gamma} := 8^{2s+1} \vee K_o(\gamma,\psi)^{2s+2\gamma+1}$. Thereby, we can use Theorem 2.5 of Tsybakov (2009), which in turn implies for any estimator S of $\mathbb{S}^s_{1/2}(L)$

$$\sup_{S\in\mathbb{S}^s_{1/2}(L)} \mathbb{P}^n_{f_Y}\big(\|\widehat{S} - S\|^2 \geq \frac{c_{\psi,\delta,\gamma}}{2} n^{-2s/(2s+2\gamma-1)}\big) \geq \frac{\sqrt{M}}{1+\sqrt{M}}\big(1 - 1/4 - \sqrt{\frac{1}{4\log(M)}}\big) \geq 0.07.$$

Note that the constant $c_{\psi,\delta,\gamma}$ does only depend on ψ, γ and δ. Hence, it is independent of the parameters s and n. The claim of Theorem 4.1.7 follows by using Markov's inequality, which completes the proof. □

Proof of Lemma 4.1.9. For any $h \in C_0^\infty(\mathbb{R})$ we can state that $\mathrm{S}[h] \in C_0^\infty(\mathbb{R})$ and thus $\mathrm{S}^j[h] \in C_0^\infty(\mathbb{R})$ for any $j \in \mathbb{N}$. Further, for $h \in C_0^\infty(\mathbb{R})$, $\int_{-\infty}^{\infty} h^{(1)}(x)d\lambda(x) = 0$, which implies that for any $\delta > 0$ and $\boldsymbol{\theta} \in \{0,1\}^K$ we have $\int_{\mathbb{R}_+} f_{\boldsymbol{\theta}}(x)d\lambda(x) = 1$.
Now due to the construction (4.9) of the functions $\psi_{k,K}$ we easily see that the function $\psi_{k,K}$ has support on $[1 + k/K, 1 + (k+1)/K]$ which leads to $\psi_{k,K}$ and $\psi_{l,K}$ having disjoint supports if $k \neq l$. Here, we want to emphasize that $\mathrm{supp}(\mathrm{S}[h]) \subseteq \mathrm{supp}(h)$ for all $h \in C_0^\infty(\mathbb{R})$ which implies that $\psi_{k,K,\gamma}$ and $\psi_{l,K,\gamma}$ have disjoint supports if $k \neq l$, too. For $x \in [1,2]^c$ we have $f_{\boldsymbol{\theta}}(x) = \exp(-x) \geq 0$. Now let us consider the case $x \in [1,2]$. In fact there is $k_o \in [\![0, K-1]\!]$ such that $x \in [1 + k_o/K, 1 + (k_o+1)/K]$ and hence

$$S_{\boldsymbol{\theta}}(x) = S_o(x) + \theta_{k_o+1}\delta K^{-s-\gamma+1}x^{-1}\psi_{k_o,K,\gamma}(x) \geq \exp(-2) - \delta 2^\gamma C_{\gamma,\infty}c_\gamma$$

since $\|\psi_{k,K,j}\|_\infty \leq 2^j C_{j,\infty}c_j K^j$ for any $k \in \{0,\dots,K-1\}$, $s \geq 1$ and $j \in \mathbb{N}$ where $c_j := \sum_{i=1}^j c_{i,j}$. Choosing $\delta \leq \delta_o(\psi,\gamma) = \exp(-2)2^{-\gamma}(C_{\gamma,\infty}c_\gamma)^{-1}$ ensures $f_{\boldsymbol{\theta}}(x) \geq 0$ for all $x \in \mathbb{R}_+$. □

Proof of Lemma 4.1.10. Our proof starts with the observation that for $t \in \mathbb{R}$ we have $\mathcal{M}_{1/2}[S_o](t) = \Gamma(1/2+it)$. Now, by applying the Stirling formula (cf. Belomestny and Goldenshluger (2020)) we get $|\Gamma(1/2+it)| \sim \exp(-\pi/2|t|)$, $|t| \geq 2$. Thus for every $s \in \mathbb{N}$ there exists L_s such that $|S_o|_s^2 \leq L$ for all $L \geq L_s$.
Next we consider $|S_o - S_{\boldsymbol{\theta}}|_s$. Let us therefore define first $\Psi_K := \sum_{k=0}^{K-1}\theta_{k+1}\psi_{k,K}$ and $\Psi_{K,j} :=$

$\mathrm{S}^j[\Psi_K]$ for $j \in \mathbb{N}$. Then we have $|S_o - S_{\boldsymbol{\theta}}|_s^2 = \delta^2 K^{-2s-2\gamma+2} |\Psi_{K,\gamma-1}|_s^2$ where $|\,.\,|_s$ is defined in 4.6. Now since for any $j \in \mathbb{N}$, $\mathrm{supp}(\Psi_{K,j}) \subset [1,2]$, $\|\Psi_{K,j}\|_\infty < \infty$ we have that the Mellin transform of $\Psi_{K,j}$ is defined for any $c \in (0,\infty)$. By a recursive application of the integration by parts we deduce that $|\mathcal{M}_{1/2}[\Psi_{K,s+\gamma-1}](t)|^2 = (1/4+t^2)^s |\mathcal{M}_{1/2}[\Psi_{k,\gamma-1}](t)|^2$, whence

$$|\Psi_{k,\gamma-1}|_s^2 \leq C_s \int_{\mathbb{R}} |\mathcal{M}_{1/2}[\Psi_{K,s+\gamma-1}](t)|^2 d\lambda(t) = C_s \int_{\mathbb{R}_+} |\Psi_{K,s+\gamma-1}(x)|^2 d\lambda(x)$$

by the Parseval identity, see equation Proposition 2.3.5 where $C_s > 0$ is a positive constant. Since the $\psi_{k,K}$ have disjoint support for different values of $k \in [\![0, K-1]\!]$ we reason that $|\Psi_{k,\gamma-1}|_s^2 \leq C_s \sum_{k=0}^{K-1} \theta_{k+1}^2 \int_{\mathbb{R}_+} |\mathrm{S}^{\gamma-1+s}[\psi_{k,K}](x)|^2 d\lambda(x)$. Applying Jensen's inequality and that $\mathrm{supp}(\psi_{k,K}) \subset [1,2]$, we obtain

$$\begin{aligned}
|\Psi_{k,\gamma-1}|_s^2 &\leq C_{(\gamma,s)} \sum_{k=0}^{K-1} \sum_{j=1}^{\gamma+s-1} c_{j,\gamma-1+s}^2 \int_1^2 x^{2j} K^{2j} \psi^{(j)}(xK-K-k)^2 d\lambda(x) \\
&\leq C_{(\gamma,s)} K^{2(\gamma-1+s)} \sum_{k=0}^{K-1} \sum_{j=1}^{\gamma+s-1} c_{j,\gamma+s}^2 4^j C_{\psi,s,\gamma}^2 K^{-1} \leq C_{(\gamma,s)} K^{2(\gamma-1+s)}.
\end{aligned}$$

Thus, $|S_o - S_{\boldsymbol{\theta}}|_s^2 \leq C_{(s,\gamma,\delta)}$ and $|S_{\boldsymbol{\theta}}|_s^2 \leq 2(|S_o - S_{\boldsymbol{\theta}}|_s^2 + |S_o|_s^2) \leq 2(C_{(s,\gamma,\delta)} + L_s) =: L_{s,\gamma,\delta,1}$. It is sufficient to show that $\int_0^\infty x f_{\boldsymbol{\theta}}(x) dx \leq L_{s,\gamma,\delta,2}$. In fact,

$$\int_{\mathbb{R}_+} x f_{\boldsymbol{\theta}}(x) d\lambda(x) = 1 + \delta K^{-s-\gamma+1} \sum_{k=0}^{K-1} \int_{1+k/K}^{1+(k+1)/K} \psi_{k,K,\gamma}(x) d\lambda(x) \leq 1 + \delta C_\gamma$$

since $\|\psi_{k,K,\gamma}\|_\infty \leq 2^\gamma C_{\gamma,\infty} c_\gamma K^\gamma = C_\gamma K^\gamma$, cf. Proof of Lemma 4.1.9. The claim follows by choosing $L_{s,\gamma,\delta} = \max\{L_{s,\gamma,\delta,1}, L_{s,\gamma,\delta,2}\}$. □

Proof of Lemma 4.1.11.
Using the fact that the functions $(\psi_{k,K,\gamma})_{k \in [\![0,K-1]\!]}$ with different index k have disjoint supports we get

$$\begin{aligned}
\|S_{\boldsymbol{\theta}} - S_{\boldsymbol{\theta}'}\|^2 &= \delta^2 K^{-2(s+\gamma-1)} \|\sum_{k=0}^{K-1} (\theta_{k+1} - \theta'_{k+1}) \psi_{k,K,\gamma-1}\|^2 \\
&= \delta^2 K^{-2(s+\gamma-1)} \rho(\boldsymbol{\theta}, \boldsymbol{\theta}') \|\psi_{0,K,\gamma-1}\|^2
\end{aligned}$$

with $\rho(\boldsymbol{\theta}, \boldsymbol{\theta}') := \sum_{j=0}^{K-1} \mathbb{1}_{\{\boldsymbol{\theta}_{j+1} \neq \boldsymbol{\theta}'_{j+1}\}}$, the Hamming distance. Now the first claim follows by showing that $\|\psi_{0,K,\gamma-1}\|^2 \geq \frac{K^{2\gamma-3} \|\psi^{(\gamma-1)}\|^2}{2}$ for K big enough. To do so, we observe that

$$\|\psi_{0,K,\gamma-1}\|^2 = \sum_{i,j \in \{1,\dots,\gamma-1\}} c_{j,\gamma-1} c_{i,\gamma-1} \int_{\mathbb{R}_+} x^{j+i+1} \psi_{0,K}^{(j)}(x) \psi_{0,K}^{(i)}(x) d\lambda(x).$$

Defining $\Sigma := \|\psi_{0,K,\gamma-1}\|^2 - \int_{\mathbb{R}_+} (x^{\gamma-1} \psi_{0,K}^{(\gamma-1)}(x))^2 d\lambda(x)$ we see that

$$\|\psi_{0,K,\gamma-1}\|^2 = \Sigma + \int_{\mathbb{R}_+} (x^{\gamma-1} \psi_{0,K}^{(\gamma-1)}(x))^2 d\lambda(x) \geq \Sigma + K^{2\gamma-3} \|\psi^{(\gamma-1)}\|^2 \tag{4.10}$$

$$\geq \frac{K^{2\gamma-3} \|\psi^{(\gamma-1)}\|^2}{2} \tag{4.11}$$

as soon as $|\Sigma| \leq \frac{K^{2\gamma-3}\|\psi^{(\gamma-1)}\|^2}{2}$. This is obviously true as soon as $K \geq K_o(\gamma,\psi)$ and thus $\|S_{\boldsymbol{\theta}} - S_{\boldsymbol{\theta}'}\|^2 \geq \frac{\delta^2\|\psi^{(\gamma-1)}\|^2}{2}K^{-2s-1}\rho(\boldsymbol{\theta},\boldsymbol{\theta}')$ for $K \geq K_o(\psi,\gamma)$. Now we use the Varshamov-Gilbert Lemma (cf. Tsybakov (2009)) which states that for $K \geq 8$ there exists a subset $\{\boldsymbol{\theta}^{(0)},\dots,\boldsymbol{\theta}^{(M)}\}$ of $\{0,1\}^K$ with $\boldsymbol{\theta}^{(0)} = (0,\dots,0)$ such that $\rho(\boldsymbol{\theta}^{(j)},\boldsymbol{\theta}^{(k)}) \geq K/8$ for all $j,k \in \{0,\dots,M\}, j \neq k$ and $M \geq 2^{K/8}$. Therefore, $\|S_{\boldsymbol{\theta}^{(j)}} - S_{\boldsymbol{\theta}^{(l)}}\|^2 \geq \frac{\|\psi^{(\gamma-1)}\|^2\delta^2}{16}K^{-2s}$.
For the second part we have $f_o = f_{\boldsymbol{\theta}^{(0)}}$, and by using $\mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}},\widetilde{f}_o) \leq \chi^2(\widetilde{f}_{\boldsymbol{\theta}},\widetilde{f}_o) := \int_{\mathbb{R}_+}(\widetilde{f}_{\boldsymbol{\theta}}(x) - \widetilde{f}_o(x))^2/\widetilde{f}_o(x)d\lambda(x)$ it is sufficient to bound the χ^2-divergence. We notice that $\widetilde{f}_{\boldsymbol{\theta}} - \widetilde{f}_o$ has support on $(0,2)$ since $f_{\boldsymbol{\theta}} - f_o$ has support on $[1,2]$ and g has support on $[0,1]$. In fact for $y > 2$, $\widetilde{f}_{\boldsymbol{\theta}}(y) - \widetilde{f}_o(y) = \int_y^\infty (f_{\boldsymbol{\theta}} - f_o)(x)x^{-1}g(y/x)d\lambda(x) = 0$. Let $\Psi_{K,\gamma} := \sum_{k=0}^{K-1}\theta_{k+1}\psi_{k,K,\gamma} = \mathrm{S}^\gamma[\sum_{k=0}^{K-1}\theta_{k+1}\psi_{k,K}] =: \mathrm{S}^\gamma[\Psi_K]$. Now by using the compact support property and a single substitution we get

$$\chi^2(\widetilde{f}_{\boldsymbol{\theta}},\widetilde{f}_o) \leq \widetilde{f}_o(2)^{-1}\|\widetilde{f}_{\boldsymbol{\theta}} - \widetilde{f}_o\|^2 = \widetilde{f}_o(2)^{-1}\delta^2K^{-2s-2\gamma+2}\|\widetilde{x^{-1}\Psi_{K,\gamma}}\|^2.$$

Let us now consider $\|\widetilde{x^{-1}\Psi_{K,\gamma}}\|^2$. In the first step we see by application of the Parseval identity that $\|\widetilde{x^{-1}\Psi_{K,\gamma}}\|^2 = \frac{1}{2\pi}\int_{-\infty}^\infty|\mathcal{M}_{1/2}[\widetilde{x^{-1}\Psi_{K,\gamma}}](t)|^2dt$. Now for $t \in \mathbb{R}$, we see by using the multiplication theorem for Mellin transforms that $\mathcal{M}_{1/2}[\widetilde{\omega^{-1}\Psi_{K,\gamma}}](t) = \mathcal{M}_{1/2}[g](t)\cdot\mathcal{M}_{1/2}[x^{-1}\mathrm{S}^\gamma[\Psi_K]](t)$. Again we have $\mathcal{M}_{1/2}[x^{-1}\mathrm{S}^\gamma[\Psi_K]](t) = (-1/2+it)^\gamma\mathcal{M}_{-1/2}[\Psi_K](t)$. Together with assumption **[G1']** we obtain

$$\|\widetilde{x^{-1}\Psi_{K,\gamma}}\|^2 \leq \frac{C_1(g)}{2\pi}\int_{\mathbb{R}}|\mathcal{M}_{-1/2}[\Psi_K](t)|^2d\lambda(t) = C_1(g)\|x^{-1}\Psi_K\|^2 \leq C_1(g)\|\psi\|^2.$$

Since $M \geq 2^K$ we have $\mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}^{(j)}},\widetilde{f}_{\boldsymbol{\theta}^{(0)}}) \leq \frac{C_1(g)\|\psi\|^2}{\widetilde{f}_o(2)\log(2)}\delta^2\log(M)K^{-2s-2\gamma+1}$. □

Proof of Theorem 4.1.8. Let us define the nested subspaces $(U_k)_{k\in\mathbb{R}_+}$ by $U_k := \{h \in \mathbb{L}^2(\mathbb{R}_+) : \forall|t| \geq k : \mathcal{M}_{1/2}[h](t) = 0\}$. For any $h \in U_k$ we consider the empirical contrast

$$\gamma_n(h) = \|h\|^2 - 2\frac{1}{2\pi}\int_{\mathbb{R}}\widehat{\mathcal{M}}(t)\frac{\mathcal{M}_{1/2}[h](-t)}{(1/2+it)\mathcal{M}_{3/2}[g](t)}d\lambda(t) = \|h\|^2 - 2n^{-1}\sum_{j=1}^n\nu_h(Y_j)$$

with $\nu_h(Y_j) := \frac{1}{2\pi}\int_{\mathbb{R}}Y_j^{1/2+it}\frac{\mathcal{M}_{1/2}[h](-t)}{(1/2+it)\mathcal{M}_{3/2}[g](t)}d\lambda(t)$. It can easily be seen that $\widehat{S}_k = \arg\min_{h\in U_k}\gamma_n(h)$ with $\gamma_n(\widehat{S}_k) = -\|\widehat{S}_k\|^2$. For $h \in U_k$ define the centered empirical process $\bar{\nu}_h := n^{-1}\sum_{j=1}^n\nu_h(Y_j) - \langle h, S\rangle$. Then we have for $h_1,h_2 \in U_k$,

$$\gamma_n(h_1) - \gamma_n(h_2) = \|h_1 - S\|^2 - \|h_2 - S\|^2 - 2\bar{\nu}_{h_1-h_2}. \tag{4.12}$$

Since $\gamma_n(\widehat{S}_k) \leq \gamma_n(S_k)$ and by the definition of $\widehat{k}$ we have $\gamma_n(\widehat{S}_{\widehat{k}}) - \widehat{\mathrm{pen}}(\widehat{k}) \leq \gamma_n(\widehat{S}_k) - \widehat{\mathrm{pen}}(k) \leq \gamma_n(S_k) - \widehat{\mathrm{pen}}(k)$ for any $k \in \mathcal{K}_n$. Now using (4.12),

$$\|S - \widehat{S}_{\widehat{k}}\|^2 \leq \|S - S_k\|^2 + 2\bar{\nu}_{\widehat{S}_{\widehat{k}}-S_k} + \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k}).$$

First we note that $U_{k_1} \subseteq U_{k_2}$ for $k_1 \leq k_2$. Let us now denote by $a \vee b := \max\{a,b\}$ and define for all $k \in \mathcal{K}_n$ the unit balls $B_k := \{h \in U_k : \|h\| \leq 1\}$. Next we deduce from $2ab \leq a^2 + b^2$

that $2\bar{\nu}_{\widehat{S}_{\widehat{k}}-S_k} \leq 4^{-1}\|\widehat{S}_{\widehat{k}} - S_k\|^2 + 4\sup_{h\in B_{\widehat{k}\vee k}} \bar{\nu}_h^2$. Furthermore, we see that $4^{-1}\|\widehat{S}_{\widehat{k}} - S_k\|^2 \leq 2^{-1}(\|\widehat{S}_{\widehat{k}} - S\|^2 + \|S - S_k\|^2)$. Putting all the facts together and defining

$$p(\widehat{k} \vee k) := 12\sigma_Y \Delta_g(\widehat{k} \vee k)n^{-1} \tag{4.13}$$

we have

$$\|S - \widehat{S}_{\widehat{k}}\|^2 \leq 3\|S - S_k\|^2 + 8\big(\sup_{h\in B_{\widehat{k}\vee k}} \bar{\nu}_h^2 - p(k \vee \widehat{k})\big)_+ + 8p(\widehat{k} \vee k) + 2\widehat{\mathrm{pen}}(k) - 2\widehat{\mathrm{pen}}(\widehat{k})$$

Assuming that $\chi \geq 48$, $4p(\widehat{k} \vee k) \leq \mathrm{pen}(k) + \mathrm{pen}(\widehat{k})$. Thus,

$$\begin{aligned}\|S - \widehat{S}_{\widehat{k}}\|^2 \leq& 6\big(\|S - S_k\|^2 + \mathrm{pen}(k)\big) + 8\max_{k\in\mathcal{K}_n}\big(\sup_{h\in B_k} \bar{\nu}_h^2 - p(k)\big)_+ \\ &+ 2(\widehat{\mathrm{pen}}(k) - 2\mathrm{pen}(k)) + 2(\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+.\end{aligned}$$

We will use the following Lemmata which will be proven afterwards.

Lemma 4.1.12:
Under the assumption of Theorem 4.1.8 we have

$$\mathbb{E}^n_{f_Y}\left(\big(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k} \bar{\nu}_h^2 - p(k)\big)_+\right) \leq \frac{C(g, \sigma, \mathbb{E}_{f_Y}(Y_1^{5/2}))}{n}$$

Lemma 4.1.13:
Under the assumption of Theorem 4.1.8 we have

$$\mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+) \leq 4\chi\frac{\mathbb{E}_{f_Y}(Y_1^2)}{\sigma_Y n}.$$

Applying the lemmata and using the fact that $\mathbb{E}^n_{f_Y}(\widehat{\mathrm{pen}}(k)) = 2\mathrm{pen}(k)$,

$$\mathbb{E}^n_{f_Y}(\|S - \widehat{S}_{\widehat{k}}\|^2) \leq 6\big(\|S - S_k\|^2 + \mathrm{pen}(k)\big) + \frac{C(g, \chi, f)}{n}.$$

Since this inequality holds for all $k \in \mathcal{K}_n$, it implies the claim. □

Proof of Lemma 4.1.12. For the first summand we see

$$\begin{aligned}\mathbb{E}^n_{f_Y}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k} \bar{\nu}_h^2 - p(k))_+) \leq& 2\mathbb{E}^n_{f_Y}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k} |\bar{\nu}_{h,1}|^2 - \frac{1}{2}p(k))_+) + 2\mathbb{E}^n_{f_Y}(\max_{k\in\mathcal{K}_n}\sup_{h\in B_k} |\bar{\nu}_{h,2}|^2),\\ =:& M_1 + M_2\end{aligned}$$

where we split the process for some sequence $(d_n)_{n\in\mathbb{N}}$ in the following way

$$\begin{aligned}\bar{\nu}_{h,1} :=& n^{-1}\sum_{j=1}^n \nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{1/2}) - \mathbb{E}_{f_Y}(\nu_h(Y_1)\mathbb{1}_{(0,d_n)}(Y_1^{1/2}))\\ \text{and } \bar{\nu}_{h,2} :=& n^{-1}\sum_{j=1}^n \nu_h(Y_j)\mathbb{1}_{dc_n,\infty)}(Y_j^{1/2}) - \mathbb{E}_{f_Y}(\nu_h(Y_1)\mathbb{1}_{(d_n,\infty)}(Y_1^{1/2})).\end{aligned}$$

To bound M_1 twe will use the Talagrand inequality 3.10 on $\mathbb{E}^n_{f_Y}\left((\sup_{h\in B_k}|\bar\nu_{h,1}|^2-\frac{1}{2}p(k))_+\right)$. Indeed, we have

$$M_1 \le \sum_{k=1}^{K_n} \mathbb{E}^n_{f_Y}(\sup_{h\in B_k}|\bar\nu_{h,1}|^2-\frac{1}{2}p(k))_+,$$

which will be used to show the claim. We want to emphasize that we are able to apply the Talagrand inequality on the sets B_k since each of them has a dense countable subset and due to continuity arguments. In order to apply Talagrand's inequality, we need to find the constants Ψ,ψ,τ such that

$$\sup_{h\in B_k}\sup_{y>0}|\nu_h(y)\mathbb{1}_{(0,d_n)}(y^{1/2})|\le\psi;\quad \mathbb{E}^n_{f_Y}(\sup_{h\in B_k}|\bar\nu_{h,1}|)\le\Psi;$$
$$\sup_{h\in B_k}\frac{1}{n}\sum_{j=1}^n \mathbb{V}\mathrm{ar}_{f_Y}(\nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{1/2}))\le\tau.$$

We start to determine the constant Ψ^2. Let us define $\widetilde{\mathcal{M}}(t):=n^{-1}\sum_{j=1}^n Y_j^{1/2+it}\mathbb{1}_{(0,d_n)}(Y_j^{1/2})$ as an unbiased estimator of $\mathcal{M}_{3/2}[f_Y\mathbb{1}_{(0,d_n)}](t)$ and

$$\widetilde{S}_k(x):=\frac{1}{2\pi}\int_{-k}^{k}x^{-1/2-it}\frac{\widetilde{\mathcal{M}}(t)}{(1/2+it)\mathcal{M}_{3/2}[g](t)}d\lambda(t)$$

where $n^{-1}\sum_{j=1}^n\nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j)=\langle\widetilde{S}_k,h\rangle$. Thus, we have for any $h\in B_k$ that $\bar\nu_{h,1}^2=\langle h,\widetilde{S}_k-\mathbb{E}^n_{f_Y}(\widetilde{S}_k)\rangle^2\le\|h\|^2\|\widetilde{S}_k-\mathbb{E}^n_{f_Y}(\widetilde{S}_k)\|^2$. Since $\|h\|\le 1$, we get

$$\mathbb{E}^n_{f_Y}(\sup_{h\in B_k}\bar\nu_{h,1}^2)\le\mathbb{E}^n_{f_Y}(\|\widetilde{S}_k-\mathbb{E}^n_{f_Y}(\widetilde{S}_k)\|^2)=\frac{1}{2\pi}\int_{-k}^{k}\frac{\mathbb{V}\mathrm{ar}^n_{f_Y}(\widetilde{\mathcal{M}}(t))}{(1/4+t^2)|\mathcal{M}_{3/2}[g](t)|^2}d\lambda(t)$$

which implies

$$\mathbb{E}^n_{f_Y}(\sup_{h\in B_k}\bar\nu_{h,1}^2)\le\sigma_Y\frac{\Delta_g(k)}{n}=:\Psi^2.$$

Thus $6\Psi^2=\frac{1}{2}p(k)$.
Next we consider ψ. Let $y>0$ and $h\in B_k$. Then using the Cauchy-Schwarz inequality,

$$\begin{aligned}|\nu_h(y)\mathbb{1}_{(0,d_n)}(y)|^2&=(2\pi)^{-2}d_n^2\Big|\int_{-k}^{k}y^{it}\frac{\mathcal{M}_{1/2}[h](-t)}{(1/2+it)\mathcal{M}_{3/2}[g](t)}d\lambda(t)\Big|^2\\&\le(2\pi)^{-1}d_n^2\int_{-k}^{k}|(1/2+it)\mathcal{M}_{3/2}[g](t)|^{-2}d\lambda(t)\le d_n^2\Delta_g(k)=:\psi^2\end{aligned}$$

since $|y^{it}|=1$ for all $t\in\mathbb{R}$.
Next we consider τ. In fact for $h\in B_k$ we can conclude

$$\mathbb{V}\mathrm{ar}^n_{f_Y}(\nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{1/2}))\le\mathbb{E}^n_{f_Y}(\nu_h(Y_j)^2)\le\|f_Yx^1\|_\infty\|\nu_h\|^2_{x^{-1}}\le\|gx\|_\infty\mathbb{E}_f(X_1)\|\nu_h\|^2_{x^{-1}}.$$

Applying the Plancherel identity delivers

$$\mathbb{V}\mathrm{ar}_{f_Y}(\nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{1/2}))\le\mathbb{E}_f(X_1)\|xg\|_\infty\frac{1}{2\pi}\int_{-k}^{k}\frac{|\mathcal{M}_{1/2}[h](t)|^2}{|(1/2+it)\mathcal{M}_{3/2}[g](t)|^2}d\lambda(t).$$

Now since $\|h\|^2 \le 1$, and for $G_k(t) := \mathbb{1}_{[-k,k]}(t)|(1/2+it)\mathcal{M}_{3/2}[g](t)|^{-2}$,

$$\sup_{h\in B_k} \frac{1}{n}\sum_{j=1}^{n} \mathbb{V}\mathrm{ar}_{f_Y}(\nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{1/2})) \le \sigma_X \|G_k\|_\infty \|xg\|_\infty =: \tau,$$

where $\sigma_X = \mathbb{E}_f(X_1)$. Hence, with $\sigma_U := \mathbb{E}_g(U_1)$ we have $\frac{n\Psi^2}{6\tau} = \frac{\sigma_U \Delta_g(k)}{6\|xg\|_\infty \|G_k\|_\infty}$ and $\frac{n\Psi}{100\psi} = \frac{\sqrt{\sigma_Y n}}{100 d_n}$. Now, choosing $d_n := \frac{\sqrt{\sigma_Y n}}{a100\log(n)}$ gives $\frac{n\Psi}{100\psi} = a\log(n)$, and we deduce

$$\begin{aligned}\mathbb{E}^n_{f_Y}\left(\big(\sup_{h\in B_k} \bar{\nu}^2_{h,1} - \frac{1}{2}p(k)\big)_+ \right) \le \frac{C}{n}\Bigg(&\sigma_X\|G_k\|_\infty\|xg\|_\infty \exp\left(-\frac{\pi\sigma_U\Delta_g(k)}{3\|xg\|_\infty\|G_k\|_\infty}\right) \\ &+\frac{\sigma_Y\Delta_g(k)}{\log(n)^2} n^{-a}\Bigg).\end{aligned}$$

Under **[G1]** we have $C_g k^{2\gamma-1} \ge \Delta_g(k) \ge c_g k^{2\gamma-1}$ and for all $t\in\mathbb{R}$ it holds true that $c_g k^{2\gamma-2} \le |G_k(t)| \le C_g k^{2\gamma-2}$. Hence,

$$\sum_{k=1}^{K_n} \mathbb{E}^n_{f_Y}\left(\big(\sup_{h\in B_k} \bar{\nu}^2_{h,1} - \frac{1}{2}p(k)\big)_+ \right) \le \frac{C\sigma_X}{n}\left(\sum_{k=1}^{K_n} C(g)k^{2\gamma-1}\exp(-C(g)k) + \sum_{k=1}^{K_n} C(g)\frac{k^{2\gamma-1}}{\log(n)^2 n^a}\right)$$

where the first sum is bounded in $n\in\mathbb{N}$. The second sum can be bounded by $C(g)n^{\frac{2\gamma}{2\gamma-1}-a}/\log(n)^2$ which by choosing $a = \frac{2\gamma}{2\gamma-1}$ ensures the boundedness in $n\in\mathbb{N}$. Thus we have

$$\sum_{k=1}^{K_n} \mathbb{E}^n_{f_Y}\left(\big(\sup_{h\in B_k} \bar{\nu}^2_{h,1} - \frac{1}{2}p(k)\big)_+ \right) \le \frac{C(g)\sigma_X}{n}.$$

Now, we consider M_2. Let us define $\overline{S}_k := \widehat{S}_k - \widetilde{S}_k$. Then from $\nu_{h,2} = \nu_h - \nu_{h,1}$ we deduce $\nu^2_{h,2} = \langle \overline{S}_k - \mathbb{E}^n_{f_Y}(\overline{S}_k), h\rangle^2 \le \|\overline{S_k} - \mathbb{E}^n_{f_Y}(\overline{S}_k)\|^2$ for any $h\in B_k$. Further, for any $k\in\mathcal{K}_n$, $\|\overline{S_k} - \mathbb{E}^n_{f_Y}(\overline{S}_k)\|^2 \le \|\overline{S}_{K_n} - \mathbb{E}^n_{f_Y}(\overline{S}_{K_n})\|^2$ and

$$\begin{aligned}\mathbb{E}(\|\overline{S}_{K_n} - \mathbb{E}(\overline{S}_{K_n}))\|^2) &= \frac{1}{2\pi}\int_{-K_n}^{K_n} \frac{\mathbb{V}\mathrm{ar}(\widehat{\mathcal{M}}(t) - \widetilde{\mathcal{M}}(t))}{|(1/2+it)\mathcal{M}_{3/2}[g](t)|^2} d\lambda(t) \\ &\le \frac{\Delta_g(K_n)}{n}\mathbb{E}_{f_Y}(Y_1\mathbb{1}_{(d_n,\infty)}(Y_1^{1/2})).\end{aligned}$$

Then, we have by the definition of K_n that for any $p\in\mathbb{N}$ holds

$$\mathbb{E}^n_{f_Y}(\|\overline{S}_{K_n} - \mathbb{E}^n_{f_Y}(\overline{S}_{K_n}))\|^2) \le d_n^{-p}\mathbb{E}_{f_Y}(Y_1^{1+p/2}) \le \frac{C(g,\sigma_X,\mathbb{E}_f(X_1^{5/2}))}{n}$$

choosing $p=3$. These two bounds imply

$$\mathbb{E}^n_{f_Y}\left(\max_{k\in\mathcal{K}_n}\left(\sup_{h\in B_k} \overline{\nu}^2_h - p(k)\right)_+ \right) \le \frac{C(g,\sigma_X,\mathbb{E}_f(X_1^{5/2})}{n}.$$

$\square$

Proof of Lemma 4.1.13. First, $\mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+) = 2\chi\mathbb{E}^n_{f_Y}((\sigma_Y/2 - \widehat{\sigma}_Y)_+\Delta_g(\widehat{k})n^{-1}) \leq 2\chi\mathbb{E}^n_{f_Y}((\sigma_Y/2 - \widehat{\sigma}_Y)_+)$. On $\Omega_Y := \{|\sigma_Y - \widehat{\sigma}_Y| \leq \sigma_Y/2\}$ we have $\sigma_Y/2 - \widehat{\sigma}_Y \leq 0$. Therefore,

$$\mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+) \leq 2\chi\mathbb{E}^n_{f_Y}((\sigma_Y/2 - \widehat{\sigma}_Y)_+\mathbb{1}_{\Omega^c}) \leq 2\chi\sqrt{\mathbb{V}\mathrm{ar}^n_{f_Y}(\widehat{\sigma}_Y)\mathbb{P}^n_{f_Y}(\Omega^c)}$$

by applying the Cauchy-Schwarz inequality. Next, by Markov's inequality, $\mathbb{P}^n_{f_Y}[|\widehat{\sigma}_Y - \sigma_Y| \geq \sigma_Y/2] \leq 4\mathbb{V}\mathrm{ar}^n_{f_Y}(\widehat{\sigma}_Y)\sigma_Y^{-2}$ which implies

$$\mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+) \leq 4\chi\mathbb{V}\mathrm{ar}^n_{f_Y}(\widehat{\sigma}_Y)\sigma_Y^{-1} \leq 4\chi\mathbb{E}^n_{f_Y}(Y_1^2)\sigma_Y^{-1}n^{-1}.$$

□

4.2 Under multiplicative measurement errors for stationary processes

4.2.1 Introduction

In this section, we study again the estimation of the unknown survival function $S : \mathbb{R}_+ \to [0,1]$ of a positive random variable X given a sample

$$Y_j = X_j U_j, \quad j \in [\![n]\!]$$

where $(U_j)_{j\in[\![n]\!]}$ is an i.i.d. sample independent of $(X_j)_{j\in[\![n]\!]}$. In contrary to Section 4.1, we will only assume that $(X_j)_{j\in[\![n]\!]}$ is strictly stationary, compare Definition 3.6.1, and study the behavior of the survival function estimator $(\widehat{S}_k)_{k\in\mathbb{R}_+}$ presented in Equation (4.4) where

$$\widehat{S}_k(x) := \frac{1}{2\pi}\int_{(-k,k)} x^{-1/2-it}\frac{\widehat{\mathcal{M}}_Y(t)}{(1/2+it)\mathcal{M}_{3/2}[g](t)}d\lambda(t), \quad x \in \mathbb{R}_+,$$

where $\widehat{\mathcal{M}}_Y(t) := n^{-1}\sum_{j\in[\![n]\!]} Y_j^{1/2+it}, t \in \mathbb{R}$.
We will consider a different characterization of dependence than in Section 3.6, the so-called Bernoulli shift processes. We introduce the so-called functional dependence measure which is defined for Bernoulli shift processes. This wide class of processes will be presented first. The following definition of Bernoulli shift is based on Dedecker et al. (2007).

Definition 4.2.1 (Bernoulli shift process)**:**
Let $H : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ be a measurable function. Let $(\xi_j)_{j\in\mathbb{Z}}$ be a strictly stationary sequence of real-valued random variables. A BERNOULLI SHIFT with innovation process $(\xi_n)_{n\in\mathbb{Z}}$ is defined as

$$X_j := H((\xi_{j-k}))_{k\in\mathbb{Z}}, \quad j \in \mathbb{Z}.$$

This sequence is strictly stationary.

While the Definition 4.2.1 from Dedecker et al. (2007) is more general, we will mainly focus on the case where $(\xi_j)_{j\in\mathbb{Z}}$ is an i.i.d. sequence and $H : \mathbb{R}^{\mathbb{N}_0} \to \mathbb{R}$, in other words X_j only depends on $(\xi_{j-k})_{k\in\mathbb{N}_0}$. One example of such a process is presented next.

Example 4.2.2 (AR(1)-process)**:**
Let $(\xi_j)_{j\in\mathbb{Z}}$ be an i.i.d. sequence and $\rho \in \mathbb{R}$ such that $|\rho| < 1$ and $\mathbb{E}(|\xi_1|^p) < \infty$ for $p \geq 1$. Then the process

$$X_n := \sum_{j=0}^{\infty}\rho^j\xi_{n-j}, \quad n \in \mathbb{N}_0,$$

is called a AR(1)-process. Here, the infinite sum is defined with respect to the $\mathbb{E}(|\cdot|^p)^{1/p}$-norm. This process is a Bernoulli shift process. In the literature, this process is commonly expressed through the recursion scheme

$$X_n = \rho X_{n-1} + \xi_n, \quad n \in \mathbb{N}_0.$$

For an independent copy ξ^*_{-k} of ξ_{-k}, $k \in \mathbb{N}$, we define

$$X_0^{*(-k)} := (\xi_0, \xi_{-1}, \dots, \xi_{-k+1}, \xi^*_{-k}, \xi_{-k-1}, \dots)$$

which is needed for the following definition.

Definition 4.2.3 (Functional Dependence measure)**:**
Let $(X_j)_{j\in\mathbb{Z}}$ be a Bernoulli shift process with $\mathbb{E}(|X_0|^p) < \infty$ for $p \in \mathbb{R}_+$. Then, we define for any $p \in \mathbb{R}_+$ the FUNCTIONAL DEPENDENCE MEASURE

$$\delta_p(k) := \mathbb{E}(|X_0 - X_0^{*(-k)}|^p)^{1/p}, \quad k \in \mathbb{N}_0.$$

For AR(1) processes, these coefficients can easily be bounded.

Example 4.2.4 (Continuation of Example 4.2.2)**:**
Let $\mathbb{E}(|\xi_0|^p) < \infty$ for $p \in \mathbb{R}_+$. Then for any $k \in \mathbb{N}$ we have

$$\delta_p(k) = \mathbb{E}(|\rho^k(\xi_{-k} - \xi^*_{-k})|^p)^{1/p} \leq 2|\rho|^k \mathbb{E}(|\xi_0|^p)^{1/p}.$$

As a consequence we deduce that if the p-th moment of the innovation ξ_0 exists, then we can state that the sequence $(\delta_p(k))_{k\in\mathbb{N}}$ is geometrically decreasing.

Before studying the risk of our estimator, let us state the following results for exponential of AR(1) processes. Its proof is postponed to the appendix.

Proposition 4.2.5 (Exponential of AR(1)-processes)**:**
Let $(X_k)_{k\in\mathbb{N}}$ be an AR(1)-process with $|\rho| < 1$ and increment sequence $(\xi_j)_{j\in\mathbb{Z}}$. Then for $Z_j = \exp(X_j) = H((\xi_{(j-k)})_{k\in\mathbb{N}})$, $j \in \mathbb{N}_0$, and $p \in \mathbb{R}_+$ with $\mathbb{E}(|\xi_0|^{2p}), \mathbb{E}(Z_0^{2p}) < \infty$, holds

$$\delta_p(k) = \mathbb{E}(|Z_0 - Z_0^{*(-k)}|^p)^{1/p} \leq 4\rho^k \mathbb{E}(Z_0^{2p})^{1/(2p)} \mathbb{E}(|\xi_0|^{2p})^{1/(2p)}, k \in \mathbb{N}.$$

4.2.2 Estimation strategy

By the strict stationarity, we have $\mathbb{P}^{X_1} = \mathbb{P}^{X_j}$, $j \in [\![n]\!]$ implying that $\mathbb{E}^n_{f_Y}(\widehat{S}_k) = S_k$ and moreover,

$$\mathbb{E}^n_{f_Y}(\|S - \widehat{S}_k\|^2) = \|S - S_k\|^2 + \mathbb{E}^n_{f_Y}(\|\widehat{S}_k - S_k\|^2), \quad k \in \mathbb{R}_+$$

where we stay in the notation of Section 4.1. As mentioned in Section 3.6.2, we can use the conditional expectation $\mathbb{E}_{|X}(\cdot) := \mathbb{E}(\cdot|\sigma((X_j)_{j\in[\![n]\!]}))$ to decompose the variance term $\mathbb{E}^n_{f_Y}(\|\widehat{S}_k - S_k\|^2)$ into a term driven by the underlying inverse problem and a term characterized by the dependency structure of $(X_j)_{j\in[\![n]\!]}$. The proof of the following Proposition can be found in the proof section due to strong similarity to the proof of Proposition 3.6.4 in Section 3.6.2.

Proposition 4.2.6 (Upper bound of the risk)**:**
Assume that $\sigma_Y := \mathbb{E}_{f_Y}(Y_1) < \infty$ and that **[G0]** holds. Then for any $k \in \mathbb{R}_+$,

$$\mathbb{E}^n_{f_Y}(\|\widehat{S}_k - S_k\|^2) \leq \sigma_Y \frac{\Delta_g(k)}{n} + \frac{1}{2\pi}\int_{(-k,k)} \frac{\mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t))}{1/4+t^2} d\lambda(t)$$

where $\widehat{\mathcal{M}}_X(t) := n^{-1}\sum_{j\in[\![n]\!]} X_j^{1/2+it}$ and $\Delta_g(k) := (2\pi)^{-1}\int_{(-k,k)} |(1/2+it)\mathcal{M}_{3/2}[g(t)]|^{-2} d\lambda(t)$. Furthermore, one has

$$\mathbb{E}^n_{f_Y}(\|\widehat{S}_k - S\|^2) \leq \|S - S_k\|^2 + \sigma_Y \frac{\Delta_g(k)}{n} + \frac{1}{2\pi}\int_{(-k,k)} \frac{\mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t))}{1/4+t^2} d\lambda(t).$$

For Bernoulli shift processes, the upcoming dependence term $\mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t))$, $t \in \mathbb{R}$, in the bound of Proposition 4.2.6 can be bounded using techniques developed and proposed in Wu (2005). The rather technical proof can be found in the proof section 4.2.6.

Lemma 4.2.7 (Variance bound for Bernoulli shift processes)**:**
Let $t \in \mathbb{R}$ and $(X_j)_{j\in\mathbb{N}}$ a Bernoulli shift process with $(\delta_p(k))_{k\in\mathbb{N}}$ defined in Definition 4.2.1 then

$$(i) \quad \mathbb{V}\mathrm{ar}^n_f\left(\sum_{j\in[\![n]\!]} X_j^{1/2+it}\right) \leq 32n(1+|t|)\left(\sum_{\ell\in\mathbb{N}_0} \delta_1(\ell)^{1/2}\right)^2 \quad \text{and}$$

$$(ii) \quad \mathbb{V}\mathrm{ar}^n_f\left(\sum_{j\in[\![n]\!]} X_j\right) \leq n\left(\sum_{\ell\in\mathbb{N}_0} \delta_2(\ell)\right)^2.$$

Based on Proposition 4.2.6 and Lemma 4.2.7 we directly deduce the following Corollary whose proof is therefore omitted.

Corollary 4.2.8:
Under the assumptions of Proposition 4.2.6 and for a Bernoulli shift process $(X_j)_{j\in[\![n]\!]}$ with $\sum_{\ell\in\mathbb{N}_0} \delta_1(\ell)^{1/2} < \infty$, we have for $k \in \mathbb{R}_+$

$$\mathbb{E}^n_{f_Y}(\|\widehat{S}_k - S\|^2) \leq \|S - S_k\|^2 + \sigma_Y \frac{\Delta_g(k)}{n} + C\left(\sum_{\ell\in\mathbb{N}_0} \delta_1(\ell)^{1/2}\right)^2 \frac{\log(k)}{n}$$

where C is a numerical positive constant.

Similarly to Section 4.1, we will assume that g is a smooth error density, compare Definition 3.2.4. More precisely, we assume that there exists $\gamma > 1/2$ and positive constants c_g, C_g such that

$$\forall t \in \mathbb{R} : c_g(1+t^2)^{-\gamma/2} \leq |\mathcal{M}_{3/2}[g](t)| \leq C_g(1+t^2)^{-\gamma/2}. \qquad \textbf{([G1])}$$

Then we have as in Section 4.1.3, that $\Delta_g(k) \leq C_g k^{2\gamma-1}$ for $k \in \mathbb{R}_+$. In this situation, the choice of $k \in \mathbb{R}_+$ is non-trivial as the variance term is increasing and unbounded in $k \in \mathbb{R}_+$ while the bias term is decreasing. In the following subsection we will study the behavior of the fully-data driven estimator $\widehat{S}_{\widehat{k}}$ proposed in Section 4.1.4 in the case of a stationary sample $(X_j)_{j\in[\![n]\!]}$.

4.2.3 Data-driven method

Again, let us consider the fully data-driven choice of $k \in \mathbb{R}_+$ defined in (4.7), that is for $\chi > 0$ and $K_n := \max\{k \in [\![n]\!] : \Delta_g(k) \leq n\}$ and

$$\widehat{k} := \underset{k \in [\![K_n]\!]}{\arg\min} \; -\|\widehat{S}_k\|^2 + \widehat{\mathrm{pen}}(k) \quad \text{where} \quad \widehat{\mathrm{pen}}(k) := 2\chi\widehat{\sigma}_Y \Delta_g(k) n^{-1}$$

for $\chi > 0$. Now defining $\mathrm{pen}(k) = \chi\sigma_Y\Delta_g(k)n^{-1}$, $k \in [\![K_n]\!]$, we can show the following result.

Theorem 4.2.9 (Data-driven choice of $k \in \mathbb{R}_+$)**:**
Let g satisfy **[G1]** with $\gamma > 1/2$ and $\|xg\|_\infty < \infty$. Further let $\mathbb{E}_{f_Y}(Y_1^{5/2}) < \infty$. Then for $\chi > 96$,

$$\begin{aligned}\mathbb{E}^n_{f_Y}(\|S - \widehat{S}_{\widehat{k}}\|^2) \leq & 6 \inf_{k \in [\![K_n]\!]} \big(\|S - S_k\|^2 + \mathrm{pen}(k)\big) + \frac{C(f,g)}{n} \\ & + C(f,g)\mathbb{V}\mathrm{ar}^n_f(\widehat{\sigma}_X) + \int_{(-K_n,K_n)} \frac{\mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t))}{1/4 + t^2} d\lambda(t)\end{aligned}$$

where $C(f,g) > 0$ is a constant depending on χ, g, $\mathbb{E}_f(X_1^{5/2})$ and $\sigma_X := \mathbb{E}_f(X_1)$. Moreover, $\widehat{\sigma}_X := n^{-1}\sum_{j\in[\![n]\!]} X_j$.

Proof of Theorem 4.2.9. Identically to the proof of Theorem 4.1.8 we can show

$$\begin{aligned}\|S - \widehat{S}_{\widehat{k}}\|^2 \leq & 6\big(\|S - S_k\|^2 + \mathrm{pen}(k)\big) + 8 \max_{k \in \mathcal{K}_n} \big(\sup_{h \in B_k} \bar{\nu}_h^2 - p(k)\big)_+ \\ & + 2(\widehat{\mathrm{pen}}(k) - 2\mathrm{pen}(k)) + 2(\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+.\end{aligned}$$

where

$$\bar{\nu}_h := n^{-1}\sum_{j=1}^n \nu_h(Y_j) - \langle h, S\rangle, \quad \nu_h(Y_j) := \frac{1}{2\pi}\int_{\mathbb{R}} Y_j^{1/2+it} \frac{\mathcal{M}_{1/2}[h](-t)}{(1/2+it)\mathcal{M}_{3/2}[g](t)} d\lambda(t),$$

and

$$p(\widehat{k} \vee k) := 12\sigma_Y\Delta_g(\widehat{k} \vee k)n^{-1}.$$

We will use the following Lemmata which will be proven afterwards in the proof section 4.2.6.

Lemma 4.2.10:
Under the assumption of Theorem 4.2.9 we have

$$\begin{aligned}\mathbb{E}^n_{f_Y}\big(\max_{k\in\mathcal{K}_n}\big(\sup_{h\in B_k}\bar{\nu}_h^2 - p(k)\big)_+\big) \leq & C(g,\sigma,\mathbb{E}_{f_Y}(Y_1^{5/2}))\left(\frac{1}{n} + \mathbb{V}\mathrm{ar}^n_f(\widehat{\sigma}_X)\right) \\ & + \frac{1}{2\pi}\int_{(-K_n,K_n)} \frac{\mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t))}{1/4+t^2} d\lambda(t)\end{aligned}$$

Lemma 4.2.11:
Under the assumption of Theorem 4.2.9 we have

$$\mathbb{E}^n_{f_Y}((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+) \leq 4\chi\frac{\mathbb{E}_{f_Y}(Y_1^2)}{\sigma_Y n} + 4\chi\frac{\sigma_Y}{\sigma_X^2}\mathbb{V}\mathrm{ar}^n_f(\widehat{\sigma}_X).$$

Applying the lemmata and using the fact that $\mathbb{E}^n_{f_Y}(\widehat{\text{pen}}(k)) = 2\text{pen}(k)$,

$$\mathbb{E}^n_{f_Y}(\|S - \widehat{S}_{\widehat{k}}\|^2) \leq 6\big(\|S - S_k\|^2 + \text{pen}(k)\big) + \frac{C(g,\chi,f)}{n} + C(g,\chi,\sigma_X)\mathbb{V}\text{ar}^n_f(\widehat{\sigma}_X) + \frac{1}{2\pi}\int_{(-K_n,K_n)} \frac{\mathbb{V}\text{ar}^n_f(\widehat{\mathcal{M}}_X(t))}{1/4+t^2} d\lambda(t).$$

Since this inequality holds for all $k \in \mathcal{K}_n$, it implies the claim. □

Now, for Bernoulli shift processes we can deduce the following Corollary using Lemma 4.2.7 and Theorem 4.2.9. Due to this, its proof is omitted.

Corollary 4.2.12:
Let the assumptions of Theorem 4.2.9 be fulfilled and assume further the $\sum_{\ell\in\mathbb{N}_0} \delta_1(\ell)^{1/2}, \sum_{\ell\in\mathbb{N}_0} \delta_2(\ell) < \infty$. Then

$$\mathbb{E}^n_{f_Y}(\|S - \widehat{S}_{\widehat{k}}\|^2) \leq 6 \inf_{k\in[\![K_n]\!]} \big(\|S - S_k\|^2 + \text{pen}(k)\big) + C(f,g,\delta_1(\cdot),\delta_2(\cdot))\frac{\log(n)}{n}$$

$C(f,g,\delta_1(\cdot),\delta_2(\cdot) > 0$ is a constant depending on χ, g, $\mathbb{E}_f(X_1^{5/2})$,$\sigma_X := \mathbb{E}_f(X_1)$ and the sequences $(\delta_1(\ell))_{\ell\in\mathbb{N}_0}$ and $(\delta_2(\ell))_{\ell\in\mathbb{N}_0}$.

4.2.4 Numerical results

Let us illustrate the performance of the fully-data driven estimator $\widehat{S}_{\widehat{k}}$ defined in 4.4 and 4.7 for stationary processes $(X_j)_{j\in\mathbb{N}}$. More precisely, we will consider examples of Bernoulli shifts, which are either an AR(1)-process or the exponential of an AR(1)-process, compare Example 4.2.2. To do so, we consider the following stationary densities whose corresponding survival function will be estimated.

(*i*) Log-Gamma distribution: $f_1(x) = \frac{x^{-4}}{6}\log(x)^3\mathbb{1}_{(1,\infty)}(x)$,

(*ii*) Gamma distribution: $f_2(x) = \frac{1}{\Gamma(4)}x^3\exp(-x)\mathbb{1}_{\mathbb{R}_+}(x)$ and

(*iii*) Log-Normal distribution: $f_3(x) = (32\pi)^{-1/2}x^{-1}\exp(-\log^2(x)/32)\mathbb{1}_{\mathbb{R}_+}(x)$,

Example 4.2.13 (Continuation of Example 4.2.4):
Let $\rho \in (0,1)$.

- GAMMA DISTRIBUTION: Let $m \in \mathbb{N}$ and $\lambda \in \mathbb{R}_+$. The stationary distribution of the AR(1) process $(X_j)_{j\in\mathbb{Z}}$, where

$$X_j := \rho X_{j-1} + \varepsilon_j, \quad \text{with } \varepsilon_j | B_j \sim \Gamma_{(B_j,\lambda)} \text{ with } B_j \sim \mathrm{Bin}_{(m,1-\rho)},$$

  is given by an $\Gamma_{(m,\lambda)}$ distribution, as shown in Gaver and Lewis (1980).

- LOG-GAMMA DISTRIBUTION: Let $m \in \mathbb{N}$ and $\lambda \in \mathbb{R}_+$. The stationary distribution of the exponential of the AR(1) process $(X_j)_{j\in\mathbb{Z}}$, that is $X_j = \exp(\mathcal{X}_j)$ where

$$\mathcal{X}_j := \rho \mathcal{X}_{j-1} + \varepsilon_j, \quad \text{with } \varepsilon_j | B_j \sim \Gamma_{(B_j,\lambda)} \text{ with } B_j \sim \mathrm{Bin}_{(m,1-\rho)},$$

  is given by a $\mathrm{LG}_{(m,\lambda)}$ distribution.

- LOG-NORMAL DISTRIBUTION: Let $\sigma^2 \in \mathbb{R}_+$. For the AR(1) process $(\mathcal{X}_j)_{j\in\mathbb{Z}}$, where for

$$\mathcal{X}_j := \rho \mathcal{X}_{j-1} + \varepsilon_j, \quad \text{with } \varepsilon_j \sim \mathrm{N}_{(0,\sigma^2(1-\rho^2))},$$

  the stationary distribution is given by a $\mathrm{N}_{(0,\sigma^2)}$ distribution. This implies that the Bernoulli shift process $X_j := \exp(\mathcal{X}_j)$ has a stationary distribution given by a $\mathrm{LN}_{(0,\sigma^2)}$-distribution.

Exploiting Example 4.2.4 and Proposition 4.2.5, we see that all conditions are fulfilled which ensures that the sequences δ_1, respectively δ_2 are of geometrical decay, that is for $\rho \in (0,1)$ we have

$$\delta_1(k) \vee \delta_2(k) \le C_X \rho^k, k \in \mathbb{N}_0,$$

where $C_X > 0$ is a positive constant only depending on the Bernoulli shift process $(X_j)_{j\in\mathbb{Z}}$. It is worth stressing out, that for $\rho = 0$ we would find ourselves in the i.i.d. case while for $\rho = 1$ we have $X_j = X_0$ for all $j \in \mathbb{Z}$. In some sense, we can interpret ρ as an indicator of the underlying dependency. For the distribution of the error density, we consider the following three cases

a) UNIFORM DISTRIBUTION: $g_1(x) = \mathbb{1}_{[0,1]}(x)$,

b) SYMMETRIC NOISE: $g_2(x) = \mathbb{1}_{(1/2,3/2)}(x)$ and

c) BETA DISTRIBUTION: $g_3(x) = 2(1-x)\mathbb{1}_{(0,1)}(x)$.

We see that g_1 and g_2 fulfill **[G1]** with the parameter $\gamma = 1$ and g_3 fulfills it with $\gamma = 2$. Due to the fact that for the true survival function holds $S(x) \in [0,1], x \in \mathbb{R}$, we can improve the estimator $\widehat{S}_{\widehat{k}}$ by defining

$$\widetilde{S}_{\widehat{k}}(x) := \begin{cases} 0 & , \widehat{S}_{\widehat{k}}(x) \le 0; \\ \widehat{S}_{\widehat{k}}(x) & , \widehat{S}_{\widehat{k}}(x) \in [0,1]; \\ 1 & , \widehat{S}_{\widehat{k}}(x) \ge 1. \end{cases}$$

By minimizing an integrated squared error over a family of histogram densities with randomly drawn partitions and weights we select $\gamma = 1$ and $\chi = 0.8$ for $\widehat{f}_{\widehat{k}}$. For the case $c)$ we choose $\chi = 0.4$.

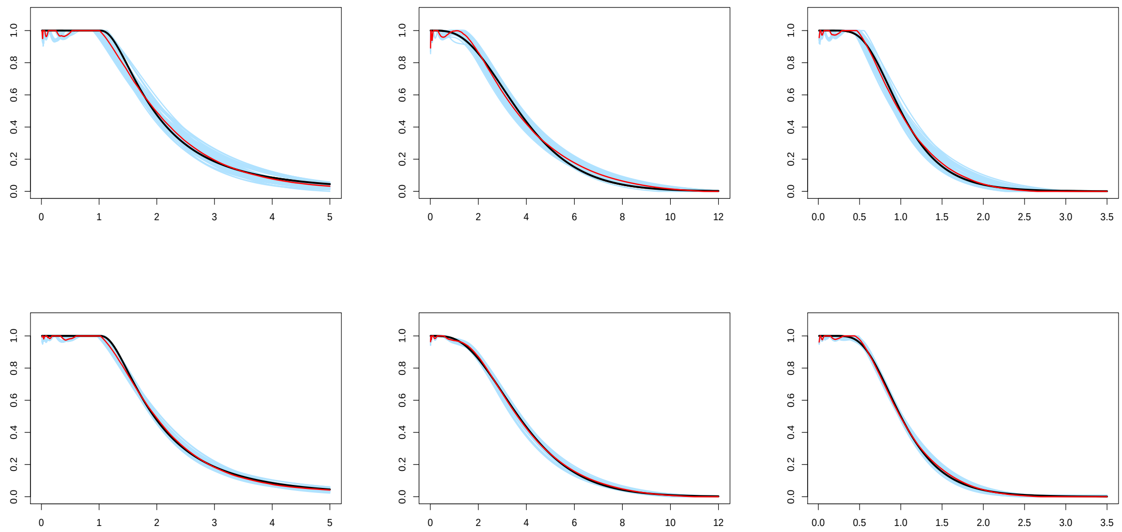

Figure 4.4: Considering the estimators $\widetilde{S}_{\widehat{k}}$, we depict 50 Monte-Carlo simulations with varying sample size $n = 500$ (top), $n = 2000$ (bottom) and in the case (i) (left), (ii) middle and (iii) (right) with error density b) and $\rho = 0.5$. The true survival function S is given by the black curve while the red curve is the point-wise empirical median of the 50 estimates.

Figure 4.4 visualizes the performance of the proposed estimator for the different examples of stationary distributions. Furthermore, it is clear to see that a higher sample size leads to a smaller integrated squared error.

In Figure 4.5, we see the different influence of varying examples of error densities. While in the left and the middle column the estimators seems to behave similar, the right column differs strongly from the other two. This observation is consistent with the risk bounds developed in Proposition 4.2.6.

Varying the value $\rho \in \{0.1, 0.5, 0.9\}$, we see in Figure 4.6 the influence of the underlying dependency structure on the behavior of the estimator.

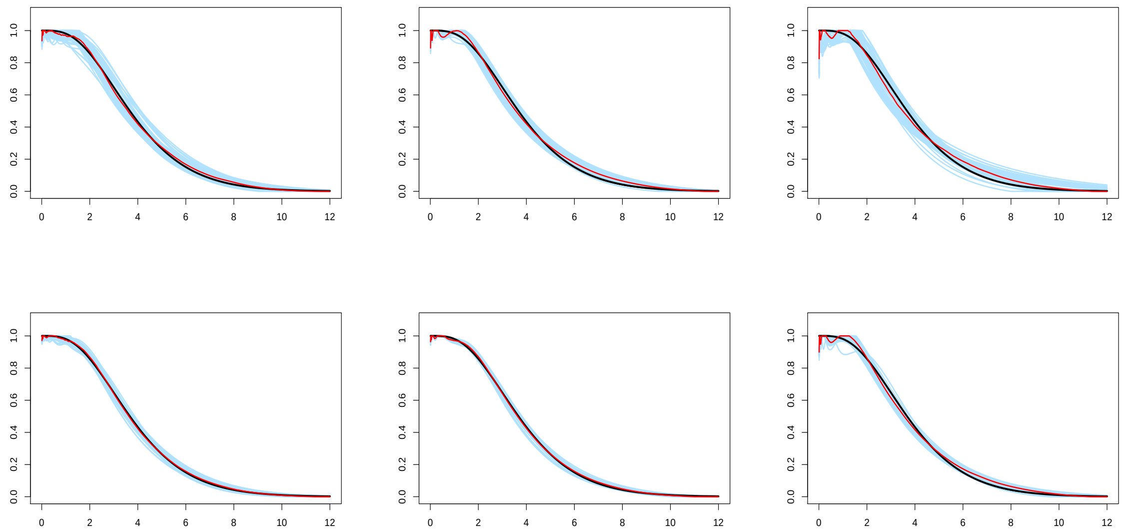

Figure 4.5: Considering the estimators $\widetilde{S}_{\widehat{k}}$, we depict 50 Monte-Carlo simulations with varying sample size $n = 500$ (top), $n = 2000$ (bottom) and in the case (ii) with varying error density a) (left), b) (middle) and c) (right) and $\rho = 0.5$. The true survival function S is given by the black curve while the red curve is the point-wise empirical median of the 50 estimates.

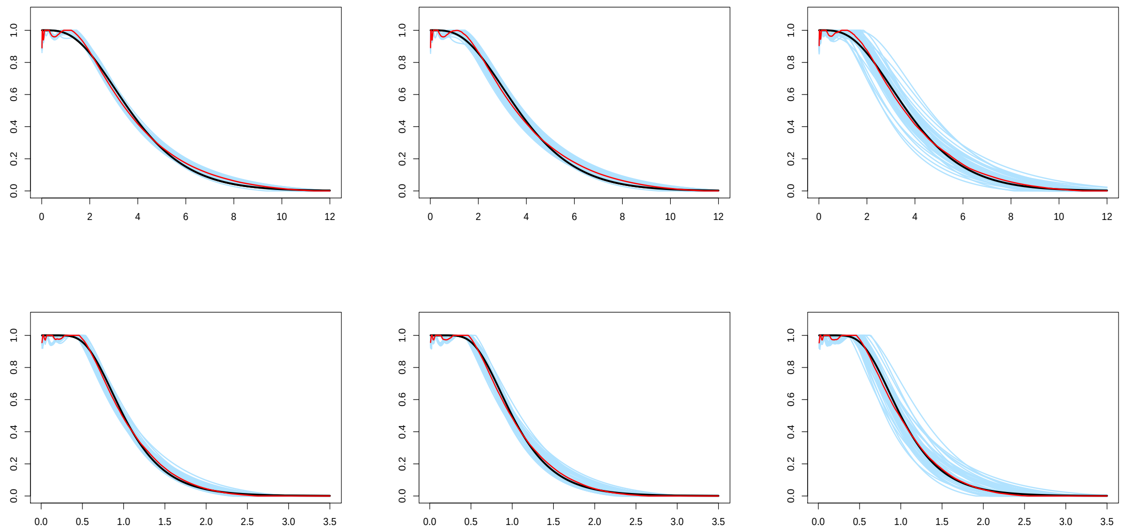

Figure 4.6: Considering the estimators $\widetilde{S}_{\widehat{k}}$, we depict 50 Monte-Carlo simulations with fixed sample size $n = 500$ in the case (ii) (top) and (iii) (bottom) with error density (b) and varying coefficient $\rho = 0.1$ (left), $\rho = 0.5$ and $\rho = 0.9$ (right). The true survival function S is given by the black curve while the red curve is the point-wise empirical median of the 50 estimates.

4.2.5 Conclusion

In this section, we studied the family of estimators $(\widehat{S}_k)_{k\in\mathbb{R}_+}$, proposed in Section 4.1 for the survival function $S:\mathbb{R}_+\to[0,1]$ of a positive random variable X based on observations $(Y_j)_{j\in[\![n]\!]}$ with multiplicative measurement error under dependency.
More precisely, we introduced the notion of Bernoulli shift processes and deduced an upper bound for the mean integrated squared error of the estimator $\widehat{S}_k$ and the fully data-driven estimator $\widehat{S}_{\widehat{k}}$. A Monte-Carlo simulation then visualize the reasonable behavior of our estimation strategy for varying scenarios as different error densities or varying dependency.

4.2.6 Proofs

Proof of Proposition 4.2.5. First we see that for all $k\in\mathbb{N}$ holds

$$\mathbb{E}(|Z_0-Z_0^{*(-k)}|^p)=\mathbb{E}\left(|\prod_{j\in\mathbb{N}_0,j\neq k}\exp(\xi_{-j})^{\rho^j}|^p|\exp(\xi_{-k})^{\rho^k}-\exp(\xi^*_{-k})^{\rho^k}|^p\right).$$

Now applying the mean-value theorem and the monotonicity of $x\mapsto\exp(\rho^k x)$ we get

$$|\exp(\xi_{-k})^{\rho^k}-\exp(\xi^*_{-k})^{\rho^k}|\leq\rho^k|\xi_{-k}-\xi^*_{-k}|\exp(\rho^k\xi_{-k}\vee\xi^*_{-k})$$

where $a\wedge b:=\min(a,b)$ for $a,b\in\mathbb{R}$. Now this implies that

$$\begin{aligned}\mathbb{E}(|Z_0-Z_0^{*(-k)}|^p)^{1/p}&\leq\rho^k\mathbb{E}((Z_0\vee Z_0^*)^p|\xi_{-k}-\xi^*_{-k}|^p)^{1/p}\\&\leq 4\rho^k\mathbb{E}(Z_0^{2p})^{1/(2p)}\mathbb{E}(|\xi_0|^{2p})^{1/(2p)}\end{aligned}$$

by application of the Cauchy-Schwarz inequality. □

Proof of Proposition 4.2.6. Due to the independence between $(U_j)_{j\in[\![n]\!]}$ and $(X_j)_{j\in[\![n]\!]}$ we derive

$$\mathbb{E}_{|X}(\widehat{\mathcal{M}}_Y(t))=\mathcal{M}_{3/2}[g](t)\widehat{\mathcal{M}}_X(t),\quad t\in\mathbb{R}.$$

By a direct calculus, we see that

$$\mathbb{E}^n_{f_Y}(|\widehat{\mathcal{M}}_Y(t)-\mathcal{M}_{3/2}[f_Y](t)|^2)=\mathbb{E}^n_{f_Y}(|\widehat{\mathcal{M}}_Y(t)-\mathbb{E}_{|X}(\widehat{\mathcal{M}}_Y(t))|^2)+|\mathcal{M}_{3/2}[g](t)|^2\mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t))$$

which implies by exploiting that

$$\mathbb{E}^n_{f_Y}(X_j^{1/2+it}\overline{X_{j'}^{1/2+it}}(U_j^{1/2+it}-\mathcal{M}_{3/2}[g](t))\overline{(U_{j'}^{1/2+it}-\mathcal{M}_{3/2}[g](t))})=\delta_{j,j'}\mathbb{E}_f(X_1)\mathbb{V}\mathrm{ar}_g(U_1^{1/2+it})$$

for $j,j'\in[\![n]\!]$ that

$$\begin{aligned}\mathbb{E}^n_{f_Y}(|\widehat{\mathcal{M}}_Y(t)-\mathcal{M}_{3/2}[f_Y](t)|^2)&=\frac{\mathbb{E}_f(X_1)\mathbb{V}\mathrm{ar}_g(U_1^{1/2+it})}{n}+|\mathcal{M}_{3/2}[g](t)|^2\mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t))\\&\leq\frac{\sigma_Y}{n}+|\mathcal{M}_{3/2}[g](t)|^2\mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t)).\end{aligned}$$

We deduce that

$$\mathbb{E}^n_{f_Y}(\|\widehat{S}_k - S_k\|^2) = \frac{1}{2\pi}\int_{(-k,k)} \frac{\mathbb{E}^n_{f_Y}(|\widehat{\mathcal{M}}_Y(t) - \mathcal{M}_{3/2}[f_Y](t)|^2)}{|(1/2+it)\mathcal{M}_{3/2}[g](t)|^2} d\lambda(t)$$
$$\leq \frac{\sigma_Y \Delta_g(k)}{n} + \frac{1}{2\pi}\int_{(-k,k)} \frac{\mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t))}{1/4+t^2} d\lambda(t).$$

The rest of the proof follows by the same steps as in the proof of Proposition 4.1.3. □

Proof of Lemma 4.2.7. We begin by proving that for all $x, y \in \mathbb{R}_+$ the inequality

$$|x^{1/2+it} - y^{1/2+it}| \leq (1+4|t|^{1/2})|x-y|^{1/2}. \tag{4.14}$$

holds true. Without loss of generality let $x > y > 0$. By the elementary inequality $|x^{1/2} - y^{1/2}| \leq |x-y|^{1/2}$ we have

$$\begin{aligned} |g(x) - g(y)| = |x^{1/2+it} - y^{1/2+it}| &\leq |x^{it}| \cdot |x-y|^{1/2} + |y|^{1/2} \cdot |x^{it} - y^{it}| \\ &\leq |x-y|^{1/2} + |y|^{1/2} \cdot |x^{it} - y^{it}| \end{aligned}$$

We then bound the second term. First we see that

$$|y|^{1/2} \cdot |x^{it} - y^{it}| \leq |y|^{1/2} \cdot (|\cos(t\log(x)) - \cos(t\log(y))| + |(\sin(t\log(y)) - \sin(t\log(y)))|)$$

Moreover,

$$\begin{aligned} |\cos(t\log(x)) - \cos(t\log(y))| &\leq |t| \cdot \log(x/y) = |t| \cdot \log\left(1 + \frac{x-y}{y}\right) \\ &\leq |t| \cdot \frac{x-y}{y}. \end{aligned}$$

where we used $\log(1+z) \leq z$ for $z > 0$. At the same time $|\cos(t\log(x)) - \cos(t\log(y))| \leq 2$. Applying both bounds together and exploiting $\min\{1, z\} \leq z^s$ for $z \geq 0$, $s \in (0,1)$,

$$\begin{aligned} |y|^{1/2} \cdot |\cos(t\log(x)) - \cos(t\log(y))| &\leq 2|y|^{1/2} \min\left\{|t|\frac{x-y}{y}, 1\right\} \\ &\leq 2|t|^{1/2}|y|^{1/2}\left|\frac{x-y}{y}\right|^{1/2} \\ &= 2|t|^{1/2}| \cdot |x-y|^{1/2}. \end{aligned}$$

A similar argument applies for the sine terms, which delivers

$$|y|^{1/2} \cdot |x^{it} - y^{it}| \leq 4|t|^{1/2} \cdot |x-y|^{1/2}$$

The case $y > x > 0$ follows analogously by interchanging the roles of x and y. Therefore, (4.14) holds true.
For the filtration $\mathcal{F}_j := \sigma(\xi_{j-i} : i \in \mathbb{N}_0)$ let us define $\mathcal{P}_{j-k}(Z) := \mathbb{E}(Z|\mathcal{F}_{j-k}) - \mathbb{E}(Z|\mathcal{F}_{j-k-1})$ for a random variable such that the increments are well-defined and let $g : \mathbb{R}_+ \to \mathbb{C}$ be a measurable function. Then by Wu (2005) the projection property of the conditional expectation and an

elementary property of δ_2 (cf. Wu (2005), Theorem 1), we have

$$\begin{aligned}\mathbb{V}\mathrm{ar}_f^n\left(\sum_{j\in[\![n]\!]} g(X_j)\right)^{1/2} &= \sum_{\ell\in\mathbb{N}_0}\mathbb{E}_f^n\left(\left(\sum_{j\in[\![n]\!]}\mathcal{P}_{j-\ell}(g(X_j))\right)^2\right)^{1/2}\\ &\le n^{1/2}\sum_{\ell\in\mathbb{N}_0}\mathbb{E}_f(|g(X_0)-g(X_0^{*(-\ell)})|^2)^{1/2}.\end{aligned}$$

Now, exploiting (4.14) we deduce (i) and (ii) since

$$\begin{aligned}\mathbb{V}\mathrm{ar}_f^n\left(\sum_{j\in[\![n]\!]} X_j^{1/2+it}\right) &\le n(1+4|t|^{1/2})^2\left(\sum_{\ell\in\mathbb{N}_0}\mathbb{E}_f(|X_0-X_0^{*(-\ell)}|)^{1/2}\right)^2\\ &= 32n(1+|t|)\left(\sum_{\ell\in\mathbb{N}_0}\delta_1(\ell)^{1/2}\right)^2\end{aligned}$$

and

$$\mathbb{V}\mathrm{ar}_f^n\left(\sum_{j\in[\![n]\!]} X_j\right) \le n\left(\sum_{\ell\in\mathbb{N}_0}\mathbb{E}_f(|X_0-X_0^{*(-\ell)}|^2)^{1/2}\right)^2 = n\left(\sum_{\ell\in\mathbb{N}_0}\delta_2(\ell)\right)^2$$

which delivers our desired statement. □

Proof of Lemma 4.2.10. The decomposition $\bar\nu_h=\bar\nu_{h,in}+\bar\nu_{h,de}$ where

$$\bar\nu_{h,in}:=n^{-1}\sum_{j=1}^n(\nu_h(Y_j)-\mathbb{E}_{|X}(\nu_h(Y_j)))\text{ and }\bar\nu_{h,de}=n^{-1}\sum_{j=1}^n\mathbb{E}_{|X}(\nu_h(Y_j)))-\mathbb{E}_{f_Y}^n(\nu_h(Y_j))$$

implies the inequality

$$\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}\bar\nu_h^2-p(k))_+\le 2\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}\bar\nu_{h,in}^2-\frac{p(k)}{2})_+ + 2\max_{k\in\mathcal{K}_n}\sup_{h\in B_k}\bar\nu_{h,de}^2.$$

Let us start with the second summand. We get for any $k\in\mathcal{K}_n$ and any $h\in B_k$,

$$\begin{aligned}\frac1n\sum_{j=1}^n\mathbb{E}_{|X}(\nu_h(Y_j))-\mathbb{E}_{f_Y}^n(\nu_h(Y_j)) &= \int_{(-k,k)}\frac{\overline{\mathcal{M}_{1/2}[h](t)}\sum_{j=1}^n\mathbb{E}_{|X}(Y_j^{1/2+it})-\mathbb{E}(Y_j^{1/2+it})}{2\pi n(1/2+it)\mathcal{M}_{3/2}[g](t)}d\lambda(t)\\ &= \int_{(-k,k)}\frac{\mathcal{M}_{1/2}[h](-t)\sum_{j=1}^n X_j^{1/2+it}-\mathbb{E}(X_j^{1/2+it})}{2\pi n(1/2+it)}d\lambda(t).\end{aligned}$$

Applying now the Cauchy-Schwarz inequality, we deduce

$$|\bar\nu_{h,de}|\le\left(\int_{(-k,k)}\frac{|n^{-1}\sum_{j=1}^n X_j^{1/2+it}-\mathbb{E}(X_j^{1/2+it})|^2}{2\pi(1/4+t^2)}d\lambda(t)\right)^{1/2}\|h\|.$$

Now since $\|h\| \leq 1$,

$$\mathbb{E}^n_{f_Y}(\max_{k\in\mathcal{K}_n}\sup_{h\in B_k}\bar{\nu}^2_{h,de}) \leq \frac{1}{2\pi}\int_{(-K_n,K_n)}\frac{\mathbb{V}\mathrm{ar}^n_f(\widehat{\mathcal{M}}_X(t))}{1/4+t^2}d\lambda(t).$$

Now, let us define $\widetilde{p}(k) := 24\sigma_U\widehat{\sigma}_X\Delta_g(k)n^{-1}$. Then,

$$\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}\bar{\nu}^2_{h,in} - \frac{1}{2}p(k))_+ = \max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}\bar{\nu}^2_{h,in} - \frac{1}{2}\widetilde{p}(k))_+ + \frac{1}{2}\max_{k\in\mathcal{K}_n}(\widetilde{p}(k) - p(k))_+.$$

For the second summand we have

$$\mathbb{E}^n_{f_Y}(\max_{k\in\mathcal{K}_n}(\widetilde{p}(k) - p(k))_+) \leq 24\sigma_U\mathbb{E}^n_{f_Y}((\widehat{\sigma}_X - \sigma_X)_+).$$

Let us define $\Omega_X := \{|\widehat{\sigma}_X - \sigma_X| \leq \sigma_X/2\}$. Then on Ω_X we have $\widehat{\sigma}_X \leq 3\sigma_X/2$ and thus $\mathbb{E}^n_{f_Y}((\widehat{\sigma}_X - \sigma_X)_+) = \mathbb{E}^n_f((\widehat{\sigma}_X - \sigma_X)_+\mathbb{1}_{\Omega_X^c}) \leq 2\sigma_X^{-1}\mathbb{V}\mathrm{ar}^n_f(\widehat{\sigma}_X)$ by application of Cauchy-Schwarz and the Markov's inequality.
For the first summand we see

$$\mathbb{E}^n_{f_Y}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}\bar{\nu}^2_{h,in} - \frac{1}{2}\widetilde{p}(k))_+) = \mathbb{E}^n_f(\mathbb{E}_{|X}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}\bar{\nu}^2_{h,in} - \frac{1}{2}\widetilde{p}(k))_+)).$$

Thus we start by considering the inner conditional expectation to bound the term. By the construction of $\bar{\nu}_{h,in}$, its summands conditioned on $\sigma(X_i, i \geq 0)$ are independent but not identically distributed. We therefore split the process for a sequence $(d_n)_{n\in\mathbb{N}}$ again in the following way

$$\begin{aligned}\bar{\nu}_{h,1} :=& n^{-1}\sum_{j=1}^n \nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{1/2}) - \mathbb{E}_{|X}(\nu_h(Y_1)\mathbb{1}_{(0,d_n)}(Y_1^{1/2}))\\ &\text{and } \bar{\nu}_{h,2} := n^{-1}\sum_{j=1}^n \nu_h(Y_j)\mathbb{1}_{(d_n,\infty)}(Y_j^{1/2}) - \mathbb{E}_{|X}(\nu_h(Y_1)\mathbb{1}_{(d_n,\infty)}(Y_1^{1/2}))\end{aligned}$$

to get

$$\begin{aligned}\mathbb{E}_{|X}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}|\bar{\nu}_{h,in}|^2 - \frac{1}{2}\widetilde{p}(k))_+) &\leq 2\mathbb{E}_{|X}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}|\bar{\nu}_{h,1}|^2 - \frac{1}{4}\widetilde{p}(k))_+) + 2\mathbb{E}_{|X}(\max_{k\in\mathcal{K}_n}\sup_{h\in B_k}|\bar{\nu}_{h,2}|^2),\\ &:= M_1 + M_2\end{aligned}$$

where we will now consider the two summands M_1, M_2 separately.
To bound M_1 we use the Talagrand inequality 3.10 on the term $\mathbb{E}_{|X}(\sup_{h\in B_k}|\bar{\nu}_{h,1}|^2 - \frac{1}{4}\widetilde{p}(k))_+$. Indeed, we have

$$M_1 \leq \sum_{k=1}^{K_n}\mathbb{E}_{|X}(\sup_{h\in B_k}|\bar{\nu}_{h,1}|^2 - \frac{1}{4}\widetilde{p}(k))_+,$$

which will be used to show the claim. We want to emphasize that we are able to apply the Talagrand inequality on the sets B_k since each set has a dense countable subset and due to continuity arguments. Further, we see that the random variables $\nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{1/2}) - \mathbb{E}_{|X}(\nu_h(Y_1)\mathbb{1}_{(0,d_n)}(Y_1^{1/2}))$,

$j = 1, \dots, n$, are conditioned on $\sigma(X_i, i \geq 0)$, centered and independent but not identically distributed. In order to apply Talagrand's inequality, we need to find the constants Ψ, ψ, τ such that

$$\sup_{h \in B_k} \sup_{y>0} |\nu_h(y) \mathbb{1}_{(0,d_n)}(y^{1/2})| \leq \psi; \quad \mathbb{E}_{|X}(\sup_{h \in B_k} |\bar{\nu}_{h,1}|) \leq \Psi;$$
$$\sup_{h \in B_k} \frac{1}{n} \sum_{j=1}^{n} \mathbb{V}\mathrm{ar}_{|X}(\nu_h(Y_j) \mathbb{1}_{(0,d_n)}(Y_j^{1/2})) \leq \tau.$$

We start to determine the constant Ψ^2. Let us define $\widetilde{\mathcal{M}}(t) := n^{-1} \sum_{j=1}^{n} Y_j^{1/2+it} \mathbb{1}_{(0,d_n)}(Y_j^{1/2})$ as an unbiased estimator of $\mathcal{M}_{3/2}[f_Y \mathbb{1}_{(0,d_n)}](t)$ and

$$\widetilde{S}_k(x) := \frac{1}{2\pi} \int_{(-k,k)} x^{-1/2-it} \frac{\widetilde{\mathcal{M}}(t)}{(1/2+it)\mathcal{M}_{3/2}[g](t)} d\lambda(t)$$

where $n^{-1} \sum_{j=1}^{n} \nu_h(Y_j) \mathbb{1}_{(0,d_n)}(Y_j) = \langle \widetilde{S}_k, h \rangle$. Thus, we have for any $h \in B_k$ that $\bar{\nu}_{h,1}^2 = \langle h, \widetilde{S}_k - \mathbb{E}_{|X}(\widetilde{S}_k) \rangle^2 \leq \|h\|^2 \|\widetilde{S}_k - \mathbb{E}_{|X}(\widetilde{S}_k)\|^2$. Since $\|h\| \leq 1$, we get

$$\mathbb{E}_{|X}(\sup_{h \in B_k} \bar{\nu}_{h,1}^2) \leq \mathbb{E}_{|X}(\|\widetilde{S}_k - \mathbb{E}_{|X}(\widetilde{S}_k)\|^2) = \frac{1}{2\pi} \int_{(-k,k)} \frac{\mathbb{E}_{|X}(|\widetilde{\mathcal{M}}(t) - \mathbb{E}_{|X}(\widetilde{\mathcal{M}}(t))|^2)}{(1/4+t^2)|\mathcal{M}_{3/2}[g](t)|^2} d\lambda(t).$$

Now since $Y_j^{1/2+it} \mathbb{1}_{(0,d_n)}(Y_j^{1/2}) - \mathbb{E}_{|X}(Y_j^{1/2+it} \mathbb{1}_{(0,d_n)}(Y_j^{1/2}))$ are independent conditioned on $\sigma(X_i : i \geq 0)$ we obtain

$$\mathbb{E}_{|X}(|\widetilde{\mathcal{M}}(t) - \mathbb{E}_{|X}(\widetilde{\mathcal{M}}(t))|^2) \leq \frac{1}{n^2} \sum_{j=1}^{n} \mathbb{E}_{|X}(Y_j \mathbb{1}_{(0,d_n)}(Y_j^{1/2})) = \frac{\sigma_U}{n} \widehat{\sigma}_X,$$

which implies

$$\mathbb{E}_{|X}(\sup_{h \in B_k} \bar{\nu}_{h,1}^2) \leq \sigma_U \widehat{\sigma}_X \frac{\Delta_g(k)}{n} =: \Psi^2.$$

Thus $6\Psi^2 = \frac{1}{4}\widetilde{p}(k)$.
Next we consider ψ. Let $y > 0$ and $h \in B_k$. Then using the Cauchy-Schwarz inequality,

$$\begin{aligned}
|\nu_h(y) \mathbb{1}_{(0,d_n)}(y)|^2 &= (2\pi)^{-2} d_n^2 | \int_{(-k,k)} y^{it} \frac{\mathcal{M}_{1/2}[h](-t)}{(1/2+it)\mathcal{M}_{3/2}[g](t)} d\lambda(t)|^2 \\
&\leq (2\pi)^{-1} d_n^2 \int_{(-k,k)} |(1/2+it)\mathcal{M}_{3/2}[g](t)|^{-2} d\lambda(t) \leq d_n^2 \Delta_g(k) =: \psi^2
\end{aligned}$$

since $|y^{it}| = 1$ for all $t \in \mathbb{R}$. Next we consider τ. In fact for $h \in B_k$ we can conclude

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}_{|X}(\nu_h(Y_j) \mathbb{1}_{(0,c_n)}(Y_j^{1/2})) \leq \mathbb{E}_{|X}(\nu_h(Y_j)^2) \\
&= \int_{-k}^{k} \int_{-k}^{k} \frac{\mathbb{E}_{|X}(Y_j^{1+i(t_1-t_2)})}{4\pi^2} \frac{\mathcal{M}_{1/2}[h](-t_1)}{(1/2+it_1)\mathcal{M}_{3/2}[g](t_1)} \frac{\mathcal{M}_{1/2}[h](t_2)}{(1/2-it_2)\mathcal{M}_{3/2}[g](-t_2)} d\lambda^2(t_1, t_2) \\
&= \int_{(-k,k)^2} \frac{X_j^{1+i(t_1-t_2)} \mathbb{E}_g(U_1^{1+i(t_1-t_2)}) \mathcal{M}_{1/2}[h](-t_1)}{4\pi^2(1/2+it_1)\mathcal{M}_{3/2}[g](t_1)} \frac{\mathcal{M}_{1/2}[h](t_2)}{(1/2-it_2)\mathcal{M}_{3/2}[g](-t_2)} d\lambda^2(t_1, t_2) \\
&= X_j \int_{\mathbb{R}_+} g(u) u \left| \mathcal{M}_{1/2}^{\dagger}[\mathbb{1}_{[-k,k]}(t) \frac{\mathcal{M}_{1/2}[h](t)}{(1/2-it)\mathcal{M}_{3/2}[g](-t)}](u) \right|^2 d\lambda(u).
\end{aligned}$$

Taking the supremum of $u \mapsto ug(u)$ and applying the Plancherel identity delivers

$$\mathbb{V}\mathrm{ar}_{|X}(\nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{1/2})) \leq X_j \|xg\|_\infty \frac{1}{2\pi}\int_{(-k,k)} \frac{|\mathcal{M}_{1/2}[h](t)|^2}{|(1/2+it)\mathcal{M}_{3/2}[g](t)|^2} d\lambda(t).$$

Now since $\|h\|^2 \leq 1$, and for $G_k(t) := \mathbb{1}_{[-k,k]}(t)|(1/2+it)\mathcal{M}_{3/2}[g](t)|^{-2}$,

$$\sup_{h\in B_k} \frac{1}{n}\sum_{j=1}^n \mathbb{V}\mathrm{ar}_{|X}(\nu_h(Y_j)\mathbb{1}_{(0,d_n)}(Y_j^{1/2})) \leq \widehat{\sigma}_X \|G_k\|_\infty \|xg\|_\infty =: \tau.$$

Hence, we have $\frac{n\Psi^2}{6\tau} = \frac{\sigma_U \Delta_g(k)}{6\|xg\|_\infty \|G_k\|_\infty}$ and $\frac{n\Psi}{100\psi} = \frac{\sqrt{\sigma_U \widehat{\sigma}_X n}}{100 d_n}$. Now, choosing $d_n := \frac{\sqrt{\sigma_U\widehat{\sigma}_X n}}{a100\log(n)}$ gives $\frac{n\Psi}{100\psi} = a\log(n)$, and we deduce

$$\mathbb{E}_{|X}\left(\big(\sup_{h\in B_k} \bar{\nu}_{h,1}^2 - \frac{1}{4}\widetilde{p}(k)\big)_+\right) \leq \frac{C}{n}\left(\widehat{\sigma}_X\|G_k\|_\infty\|xg\|_\infty \exp(-\frac{\pi\sigma_U\Delta_g(k)}{3\|xg\|_\infty\|G_k\|_\infty}) \right.$$
$$\left. + \frac{\sigma_U\widehat{\sigma}_X\Delta_g(k)}{\log(n)^2} n^{-a}\right).$$

Under **[G1]** we have $C_g k^{2\gamma-1} \geq \Delta_g(k) \geq c_g k^{2\gamma-1}$ and for all $t \in \mathbb{R}$ it holds true that $c_g k^{2\gamma-2} \leq |G_k(t)| \leq C_g k^{2\gamma-2}$. Hence,

$$\sum_{k=1}^{K_n} \mathbb{E}_{|X}\left(\big(\sup_{h\in B_k} \bar{\nu}_{h,1}^2 - \frac{1}{4}\widetilde{p}(k)\big)_+\right) \leq \frac{C\widehat{\sigma}_X}{n}\left(\sum_{k=1}^{K_n} C(g)k^{2\gamma-1}\exp(-C(g)k) + \sum_{k=1}^{K_n} C(g)\frac{k^{2\gamma-1}}{\log(n)^2 n^a}\right)$$

where the first sum is bounded in $n \in \mathbb{N}$. The second sum can be bounded by $C(g)n^{\frac{2\gamma}{2\gamma-1}-a}/\log(n)^2$ which by choosing $a = \frac{2\gamma}{2\gamma-1}$ ensures the boundedness in $n \in \mathbb{N}$. Thus we have

$$\sum_{k=1}^{K_n} \mathbb{E}_{|X}\left(\big(\sup_{h\in B_k} \bar{\nu}_{h,1}^2 - \frac{1}{4}\widetilde{p}(k)\big)_+\right) \leq \frac{C(g)\widehat{\sigma}_X}{n}.$$

Now, we consider M_2. Let us define $\overline{S}_k := \widehat{S}_k - \widetilde{S}_k$. Then from $\nu_{h,2} = \nu_{h,in} - \nu_{h,1}$ we deduce $\nu_{h,2}^2 = \langle \overline{S}_k - \mathbb{E}_{|X}(\overline{S}_k), h\rangle^2 \leq \|\overline{S_k} - \mathbb{E}_{|X}(\overline{S}_k)\|^2$ for any $h \in B_k$. Further, for any $k \in \mathcal{K}_n$, $\|\overline{S_k} - \mathbb{E}_{|X}(\overline{S}_k)\|^2 \leq \|\overline{S}_{K_n} - \mathbb{E}_{|X}(\overline{S}_{K_n})\|^2$ and

$$\mathbb{E}_{|X}(\|\overline{S}_{K_n} - \mathbb{E}_{|X}(\overline{S}_{K_n}))\|^2) = \frac{1}{2\pi}\int_{-K_n}^{K_n} \mathbb{V}\mathrm{ar}_{|X}(\widehat{\mathcal{M}}(t) - \widetilde{\mathcal{M}}(t))|(1/2+it)\mathcal{M}_{3/2}[g](t)|^{-2}dt$$
$$\leq \frac{1}{n^2}\sum_{j=1}^n \mathbb{E}_{|X}(Y_j\mathbb{1}_{(d_n,\infty)}(Y_j^{1/2}))\Delta_g(K_n).$$

Let us define the event $\Xi_X := \{\widehat{\sigma}_X \geq \sigma_X/2\}$. Then, we have

$$\frac{1}{n^2}\sum_{j=1}^n \mathbb{E}_{|X}(Y_j\mathbb{1}_{(d_n,\infty)}(Y_j^{1/2}))\Delta_g(K_n)\mathbb{1}_{\Xi_X} \leq \frac{1}{n}\sum_{j=1}^n X_j^{1+p/2}\mathbb{E}(U_j^{1+p/2})d_n^{-p}\mathbb{1}_{\Xi_X}$$

where on Ξ_X we can state that $d_n^{-p} = C(g)n^{-p/2}(\widehat{\sigma}_X)^{-p/2}\log(n)^p \leq C(g)\sigma_X^{-p/2}n^{-p/2}\log(n)^p$. Choosing $p = 3$ leads to $\mathbb{E}_{|X}(\|\overline{S}_{K_n} - \mathbb{E}_{|X}(\overline{S}_{K_n}))\|^2)\mathbb{1}_{\Xi_X} \leq \frac{C(g)\sigma_X^{-3/2}}{n}\mathbb{E}_g(U_1^{5/2})n^{-1}\sum_{j=1}^n X_j^{5/2}$. On the other hand,

$$\frac{1}{n^2}\sum_{j=1}^n \mathbb{E}_{|X}(Y_j\mathbb{1}_{(d_n,\infty)}(Y_j^{1/2}))\Delta_g(K_n)\mathbb{1}_{\Xi_X^c} \leq \frac{\sigma_X}{2}\mathbb{1}_{\Xi_X^c} \leq \frac{\sigma_X}{2}\mathbb{1}_{\Omega_X^c}.$$

These three bounds imply

$$\mathbb{E}_{f_Y}^n(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}\overline{\nu}_{h,in}^2 - \frac{1}{2}p(k))_+) \leq C(g)(\frac{\sigma_X}{2n} + \frac{\mathbb{E}_f(X_1^{5/2})}{\sigma_X^{3/2}n} + \frac{2\mathbb{V}\mathrm{ar}_f^n(\widehat{\sigma}_X)}{\sigma_X}).$$

□

Proof of Lemma 4.1.13. First we see that

$$\mathbb{E}_{f_Y}^n((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+) = 2\chi\mathbb{E}_{f_Y}^n((\sigma_Y/2 - \widehat{\sigma}_Y)_+\Delta_g(\widehat{k})n^{-1}) \leq 2\chi\mathbb{E}_{f_Y}^n((\sigma_Y/2 - \widehat{\sigma}_Y)_+).$$

On $\Omega_Y := \{|\sigma_Y - \widehat{\sigma}_Y| \leq \sigma_Y/2\}$ we have $\sigma_Y/2 - \widehat{\sigma}_Y \leq 0$. Therefore,

$$\mathbb{E}_{f_Y}^n((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+) \leq 2\chi\mathbb{E}_{f_Y}^n((\sigma_Y/2 - \widehat{\sigma}_Y)_+\mathbb{1}_{\Omega^c}) \leq 2\chi\sqrt{\mathbb{V}\mathrm{ar}_{f_Y}^n(\widehat{\sigma}_Y)\mathbb{P}_{f_Y}^n(\Omega^c)}$$

by applying the Cauchy-Schwarz inequality. Next, by Markov's inequality, $\mathbb{P}_{f_Y}^n[|\widehat{\sigma}_Y - \sigma_Y| \geq \sigma_Y/2] \leq 4\mathbb{V}\mathrm{ar}_{f_Y}(\widehat{\sigma}_Y)\sigma_Y^{-2}$ which implies $\mathbb{E}((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+) \leq 4\chi\mathbb{V}\mathrm{ar}(\widehat{\sigma}_Y)\sigma_Y^{-1}$. In analogy to the proof of Proposition 4.2.6 we get

$$\mathbb{V}\mathrm{ar}_{f_Y}^n(\widehat{\sigma}_Y) \leq \frac{\mathbb{E}_{f_Y}(Y_1^2)}{n} + \mathbb{E}_g(U_1)^2\mathbb{V}\mathrm{ar}_f^n(\widehat{\sigma}_X).$$

□

CHAPTER 5

Local estimation under multiplicative measurement errors

Contents

5.1 Introduction

In this chapter we are interested in estimating the value $\vartheta(f)$ of a known linear functional evaluated at an unknown density $f : \mathbb{R}_+ \to \mathbb{R}_+$ of a positive random variable X, when $Y = XU$ is only observable for some multiplicative positive error U. We assume that X and U are independent and that U has a known density $g : \mathbb{R}_+ \to \mathbb{R}_+$. In such a multiplicative measurement error model the density $f_Y : \mathbb{R}_+ \to \mathbb{R}_+$ (of the observable Y) is consequently given by the multiplicative convolution between f and g given through

$$f_Y(y) = (f * g)(y) = \int_{\mathbb{R}_+} f(x)g(y/x)x^{-1}d\lambda(x), \quad y \in \mathbb{R}_+,$$

compare Definition 2.4.1. The estimation of f using an independent and identically distributed (i.i.d.) sample from f_Y is therefore called a multiplicative deconvolution problem.
A list of literature concerning the statistical inference in the multiplicative measurement error model can be found in Section 3.2.1.
In this section, we focus on the estimation of the value $\vartheta(f)$ of a known linear functional of the unknown density f. The non-parametric estimation of the value of a linear functional from Gaussian white noise observations is subject of considerable literature (in case of direct observations see

Speckman (1979), Li (1982) or Ibragimov and Has'minskii (1984), while in case of indirect observations we refer to Donoho and Low (1992); Donoho (1994) or Goldenshluger and Pereverzev (2000) and references therein). In additive deconvolution linear functional estimation has been studied for instance by Butucea and Comte (2009), Mabon (2016) and Pensky (2017) to mention but a few. Notably, the estimation of the value of a linear functional evaluated at the Lévy measure assuming observations from a Lévy process has been considered by Kappus (2014). In the literature, the most studied examples for estimating linear functionals are point-wise estimation of the unknown density f, its derivatives, its associated survival, cumulative distribution function or its Laplace transform, and the estimation of their (possibly weighted) averages over a subinterval of their domains. All these examples are particular cases of our general setting. More precisely, we show below, that in each of those examples the quantity of interest can be written as the value of a linear functional evaluated at the Mellin transform $\mathcal{M}[f] : \mathbb{R} \to \mathbb{C}$ of f in the form

$$\vartheta(f) = \frac{1}{2\pi} \int_{\mathbb{R}} \Psi(-t)\mathcal{M}[f](t)d\lambda(t)$$

where $\Psi : \mathbb{R} \to \mathbb{C}$ is a known function. This particular representation allows us to use a plug-in estimator $\widehat{\vartheta}_k := \vartheta(\hat{f}_k)$ based on the density estimator $\hat{f}_k$ of f proposed in Section 3.2. It is worth noting that in Section 3.2 we have shown, that even relatively large sample sizes may not be of much help estimating accurately the density f as a whole. The reason for these poor non-parametric convergence rates is intrinsic to the considered multiplicative measurement error model as it leads in a natural way to an ill-posed inverse problem. Exploiting properties of the Mellin transform we characterize the underlying inverse problem and natural regularity conditions which borrow ideas from the inverse problems community (see e.g. Engl et al. (1996)). Those conditions essentially involve the decay of the Mellin transform of f and g. In this work we eventually impose conditions on the decay of the Mellin transform of f and g, and the decay of the function Ψ which also ensure that our estimator is well-defined. We shall access the accuracy of the proposed plug-in estimator by its mean squared error. We identify conditions on f, g and Ψ under which the plug-in estimator attains a parametric rate of convergence. We illustrate those conditions by considering different scenarios. The convergence rate of the proposed plug-in estimator $\widehat{\vartheta}_k$, however, depends crucially on a tuning parameter k, which has to be chosen non-trivially if the rate is non-parametric. Provided its optimal choice we show that uniformly over Mellin-Sobolev spaces the proposed plug-in estimator can attain minimax-optimal non-parametric rates. However, since the necessary information for an optimal choice of the tuning parameter k is widely inaccessible in practice, we propose a fully data-driven choice $\widehat{k}$. The procedure for the local selection of the tuning parameter is inspired by the work of Goldenshluger and Lepski (2011) who consider data-driven bandwidth selection in kernel density estimation. We establish an oracle inequality for the completely data-driven plug-in spectral cut-off estimator $\widehat{\vartheta}_{\widehat{k}}$ under fairly mild assumptions on f, g and Ψ. The upper bound of its mean squared error compared to the minimax rate of convergence features an additional logarithmic factor which possibly results in a deterioration of the rate. The appearance of the logarithmic factor within the rate is a known fact in the context of local estimation and it is widely considered as an acceptable price for adaptation (cf. Brown and Low (1996) in the context of non-parametric Gaussian regression or Laurent et al. (2008) given direct Gaussian observations). We shall emphasize that as usual in nonparametric estimation in comparison to the estimation of the density as a whole as only local features are of interest, it is likely that compared to the overall performance, certain local features might be estimated more accurately even with the logarithmic factor.

This Section is organized in the following way: in Section 5.2 we motivate the linear functional model by presenting several local estimation problems which can be interpreted as the estimation of the evaluation of a linear functional in f. Furthermore, we propose a plug-in estimator and prove a risk decomposition. In Section 5.3 we state an upper bound of the maximal mean squared error over Mellin-Sobolev spaces of the plug-in spectral cut-off estimator with optimally selected tuning parameter realizing a squared-bias-variance trade-off. We show lower bounds for the point-wise estimation of the unknown density f, the corresponding survival function and the derivative of f that match the upper bounds up to the constants. Thereby, we establish in this situations the minimax-optimality of the plug-in spectral cut-off estimator. We develop the data-driven plug-in estimator in Section 5.4. We state an oracle type upper bound for the mean squared error of the plug-in spectral cut-off estimator with fully-data driven choice of the tuning parameter. The proofs can be found in the Section 5.6.

5.2 Estimation strategy

In the following paragraph we introduce the linear functional. We motivate it through a collection of examples and we determine sufficient conditions to ensure that the considered objects are well-defined. We then define an estimator based on the empirical Mellin transform and the multiplication theorem for Mellin transforms. Let $c \in \mathbb{R}$ and $f \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ be fixed. In the sequel we are interested in estimating the value of the linear functional evaluated at the Mellin transform $\mathcal{M}_c[f] : \mathbb{R} \to \mathbb{C}$ of f given by

$$\vartheta(f) = \frac{1}{2\pi} \int_{\mathbb{R}} \Psi(-t) \mathcal{M}_c[f](t) d\lambda(t) \tag{5.1}$$

for a function $\Psi : \mathbb{R} \to \mathbb{C}$ with $\overline{\Psi(t)} = \Psi(-t)$ for any $t \in \mathbb{R}$. The integral in (5.1) is well-defined whenever $\Psi\mathcal{M}_c[f] \in \mathbb{L}^1(\mathbb{R})$ which we assume from now on. Obviously, the latter is trivially fulfilled, if $\Psi \in \mathbb{L}^2_{\mathbb{R}}$. However, we may emphasize that the condition $\Psi\mathcal{M}_c[f] \in \mathbb{L}^1(\mathbb{R})$ neither imposes that Ψ vanishes at infinity nor that it is square integrable. Evidently, this condition enables us depending on $\mathcal{M}_c[f]$ to deal with more demanding functionals, such as the estimation of the point-wise evaluation of the density f or its derivative. Before we present an estimator for $\vartheta(f)$ let us briefly illustrate our general approach by typical examples.

Example 5.2.1 (Point wise density evaluation)**:**
Let $x_o \in \mathbb{R}_+$. If $\mathcal{M}_c[f] \in \mathbb{L}^1_{\mathbb{R}}$ then we have

$$f(x_o) = \mathcal{M}_c^{\dagger}[\mathcal{M}_c[f]](x_o) = \frac{1}{2\pi} \int_{\mathbb{R}} x_o^{-c-it} \mathcal{M}_c[f](t) d\lambda(t) = \vartheta(f),$$

with $\Psi(t) := x_o^{-c+it}$, $t \in \mathbb{R}$, satisfying $\overline{\Psi(t)} = \Psi(-t)$ for all $t \in \mathbb{R}$.

$f \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ already implies that $|\mathcal{M}_c[f](t)| \to 0$ for $|t| \to \infty$. Thus for the following four examples, Example 5.2.2, Example 5.2.3, Example 5.2.4 and Example 5.2.5, the assumption $\Psi\mathcal{M}_c[f] \in \mathbb{L}^1(\mathbb{R})$ is trivially fulfilled due to the decay of the specific functions Ψ in each case and can be therefore omitted.

Example 5.2.2 (Point wise Cumulative distribution function evaluation)**:**
Let $x_o \in \mathbb{R}_+$. Then for $c < 1$ the function $\Psi(t) := (1-c+it)^{-1}x_o^{1-c+it}$ fulfills $\Psi\mathcal{M}_c[f] \in \mathbb{L}^1(\mathbb{R})$ and

$$\int_0^{x_o} f(x)d\lambda(x) =: F(x_o) = \frac{1}{2\pi}\int_{\mathbb{R}} x_o^{-(c-1)-it}(1-c-it)^{-1}\mathcal{M}_c[f](t)d\lambda(t) = \vartheta(f).$$

Here, $F(x_o) = \int_0^{x_o} f(x)d\lambda(x)$ is the evaluation of the c.d.f. F at the point x_o. The equality $F(x_o) = (2\pi)^{-1}\langle\Psi, \mathcal{M}_c[f]\rangle_{\mathbb{R}}$ follows from the Plancherel identity, Proposition 2.3.5, exploiting that $F(x_o) = \langle\psi, f\rangle_{x^{2c-1}}$ for $\psi(x) := x^{1-2c}\mathbb{1}_{(0,x_o)}$ with $\psi \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) \cap \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ and $\mathcal{M}_c[\psi] = \Psi$.

Example 5.2.3 (Point wise survival function evaluation)**:**
Let $x_o \in \mathbb{R}_+$. Then for $c > 1$ the function $\Psi(t) := (c-1-it)^{-1}x_o^{1-c+it}$ fulfills $\Psi\mathcal{M}_c[f] \in \mathbb{L}^1(\mathbb{R})$ and

$$\int_{x_o}^{\infty} f(x)d\lambda(x) =: S(x_o) = \frac{1}{2\pi}\int_{\mathbb{R}} x_o^{-(c-1)-it}(c-1+it)^{-1}\mathcal{M}_c[f](t)d\lambda(t) = \vartheta(f).$$

Here, $S(x_o) = \int_{x_o}^{\infty} f(x)d\lambda(x)$ is the evaluation of the survival function S at the point x_o. The equality $S(x_o) = (2\pi)^{-1}\langle\Psi, \mathcal{M}_c[f]\rangle_{\mathbb{R}}$ follows from the Plancherel identity, Proposition 2.3.5.

Example 5.2.4 (Point wise mean residual life evaluation)**:**
The MEAN RESIDUAL LIFE at the point x_o is defined by

$$m(x_o) := \mathbb{E}_f(X - x_o | X > x_o) = \frac{\mathbb{E}_f((X-x_o)\mathbb{1}_{(0,\infty)}(X-x_o))}{S(x_o)}$$

where $S(x_o) = \mathbb{P}(X > x_o)$ denotes the survival function corresponding to f, compare Guess and Proschan (1988). Interestingly, we have $\mathbb{E}_f((X-x_o)\mathbb{1}_{(x_o,\infty)}(X)) = \int_{x_0}^{\infty} S(x)d\lambda(x)$ which can be written as the value $\vartheta(f)$ of a linear functional in the form of (5.1) for $c > 2$ choosing

$$\Psi(t) = (2-c+it)^{-1}(1-c+it)^{-1}x_o^{2-c+it}.$$

Since $\vartheta(f) = \int_{x_0}^{\infty} S(x)d\lambda(x)$ is obtained by an integration of the survival function we expect a higher estimation accuracy than for the point-wise Estimation of the survival function. Consequently, the estimation of the mean residual life will roughly speaking inherent the accuracy of the estimator of its nominator.

Example 5.2.5 (Point wise Laplace transform evaluation)**:**
Consider the evaluation $L(x_o) = \int_{\mathbb{R}_+} \exp(-x_o x)f(x)d\lambda(x$ of the Laplace transform L of f at the point x_o. For $c < 1$ the function $\psi(x) := x^{1-2c}\exp(-x_o x)$, $x \in \mathbb{R}_+$ belongs to $\mathbb{L}^1(\mathbb{R}_+, x^{c-1}) \cap \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ and setting

$$\Psi(t) := \mathcal{M}_c[\psi](t) = \int_{\mathbb{R}_+} x^{-c+it}\exp(-x_o x)d\lambda(x) = x_o^{c-1-it}\Gamma(1-c+it)$$

we get $\vartheta(f) = L(x_o)$ by an application of the Plancherel identity, see Proposition 2.3.5.

Nevertheless, there we want to include to our study examples of Ψ where Ψ is not bounded. In these cases, additional assumption on the decay of $\mathcal{M}_c[f]$ are necessary. Later on, we will see that these assumptions are natural in terms of regularity assumptions on the density f to ensure that point evaluations of the density, respectively its derivative, are well-defined.

Example 5.2.6 (Point wise derivative of the density evaluation)**:**
If $t \mapsto (c+it)\mathcal{M}_c[f](t) \in \mathbb{L}^1(\mathbb{R})$ then the point evaluation $f'(x_o)$ of the derivative f' of f in x_o setting

$$\Psi(t) := -x_0^{-c-1+it}(c-it)$$

satisfies $\vartheta(f) = f'(x_o)$.

It is worth stressing out that in all five examples the quantity of interest does not depend on value of the model parameter $c \in \mathbb{R}$. However, this is not anymore true for their representations in form of (5.1), since value $\vartheta(f)$ equals a known linear functional evaluated at the Mellin transform $\mathcal{M}_c[f]$ of f. The conditions on $c \in \mathbb{R}$ in Example 5.2.1-5.2.6 and the assumption $f \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ ensure that the representation $\vartheta(f)$ is well-defined, which is essential for our estimation strategy. The value of the model parameter $c \in \mathbb{R}$ imposes amongst others moment conditions on the unobserved variables X and U. Consequently, we present the upcoming theory for almost arbitrary values of $c \in \mathbb{R}$.

Remark 5.2.7 (Relationship between survival and cumulative distribution function)**:**
Consider Example 5.2.2 and 5.2.3. Since $S = 1 - F$ there is an elementary connection between the estimation of the survival function and the estimation of the c.d.f.. For example, we eventually deduce from a c.d.f. estimator $\widehat{F}(x_o)$ a survival function estimator $\widehat{S}(x_o)$ through $\widehat{S}(x_o) := 1 - \widehat{F}(x_o)$ with same risk $\mathbb{E}_{f_Y}((\widehat{S}(x_o) - S(x_o))^2) = \mathbb{E}_{f_Y}((\widehat{F}(x_o) - F(x_o))^2)$. Thus we may define for any $c \neq 1$ a survival function (respectively c.d.f.) estimator using Example 5.2.2 and 5.2.3.

Our estimator of the quantity $\vartheta(f)$ makes uses of the convolution theorem, Proposition 2.2.5, as done in Section 3.2. To do so, let $f \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) \cap \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ and $g \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ which ensure that $\mathcal{M}_c[f_Y] = \mathcal{M}_c[f]\mathcal{M}_c[g]$. Under the mild assumption $\mathcal{M}_c[g](t) \neq 0$ for all $t \in \mathbb{R}$ we obviously have $\mathcal{M}_c[f] = \mathcal{M}_c[f_Y]/\mathcal{M}_c[g]$ allowing us to rewrite (5.1) as follows

$$\vartheta(f) = \frac{1}{2\pi}\int_{\mathbb{R}} \Psi(-t)\frac{\mathcal{M}_c[f_Y](t)}{\mathcal{M}_c[g](t)}d\lambda(t). \tag{5.2}$$

In the sequel we assume that for a fixed $c \in \mathbb{R}$ the densities f and g, and the function Ψ satisfy

$$\begin{cases} f \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) \cap \mathbb{L}^1(\mathbb{R}_+, x^{c-1}); \\ g \in \mathbb{L}^1_{\mathbb{R}_+}(x^{c-1}) \text{ with } \forall t \in \mathbb{R} : \mathcal{M}_c[g](t) \neq 0; \\ \Psi\mathcal{M}_c[f] \in \mathbb{L}^1_{\mathbb{R}} \text{ with } \forall t \in \mathbb{R} : \overline{\Psi(t)} = \Psi(-t) \end{cases} \tag{5.3}$$

which ensure that the representation of the real-valued linear functional $\vartheta(f)$ in (5.2) is well-defined.

Remark 5.2.8:
The assumption that $\mathcal{M}_c[g]$ does not vanish on $\mathbb{R}$ is rather typical and it is fulfilled by many examples of densities. Note that the density f, and therefore also $\vartheta(f)$, is without an additional assumption not identifiable anymore if $\mathcal{M}_c[g]$ vanishes for example on a non-empty interval. However, the assumption can be relaxed to oscillating or vanishing $\mathcal{M}_c[g]$ if the estimator is based on a ridge approach, compare Hall and Meister (2007) and Section 3.4. □

A naive approach based on a sample $(Y_j)_{j\in[\![n]\!]}$ from f_Y is to replace in (5.2) the quantity $\mathcal{M}_c[f_Y]$ by its empirical counterpart $\widehat{\mathcal{M}}_c(t) := n^{-1}\sum_{j\in[\![n]\!]} Y_j^{c-1+it}$, $t\in\mathbb{R}$. However, the resulting integral is not well-defined, since $\Psi\widehat{\mathcal{M}}_c/\mathcal{M}_c[g]$ is generally not integrable. We ensure integrability introducing an additional spectral cut-off regularization which leads to the following estimator

$$\widehat{\vartheta}_k := \frac{1}{2\pi}\int_{[-k,k]}\Psi(-t)\frac{\widehat{\mathcal{M}}_c(t)}{\mathcal{M}_c[g](t)}dt \quad \text{for any } k\in\mathbb{R}_+. \tag{5.4}$$

The regularization of the inverse Mellin transform using a spectral cut-off approach is inspired by Section 3.2. The following proposition establishes consistency of the estimator under a suitable choice of the cut-off parameter $k\in\mathbb{R}_+$.

Proposition 5.2.9 (Upper bound of the risk):
In addition to (5.3), assume that $\sigma_Y := \mathbb{E}_{f_Y}(Y_1^{2(c-1)}) < \infty$. Then for any $k\in\mathbb{R}_+$ we have

$$\mathbb{E}^n_{f_Y}(|\widehat{\vartheta}_k - \vartheta(f)|^2) \leq \|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|^2_{\mathbb{L}^1(\mathbb{R})} + \frac{\sigma_Y}{n}\|\mathbb{1}_{[-k,k]}\Psi/\mathcal{M}_c[g]\|^2_{\mathbb{L}^1(\mathbb{R})}. \tag{5.5}$$

If additionally $\|x^{2c-1}g\|_\infty < \infty$ setting $\Delta_{\Psi,g}(k) := (2\pi)^{-1}\|\mathbb{1}_{[-k,k]}\Psi/\mathcal{M}_c[g]\|^2_{\mathbb{R}}$ and $\mathfrak{C}_g := \|x^{2c-1}g\|_\infty/\mathbb{E}_g(U_1^{2(c-1)})$ we obtain

$$\mathbb{E}^n_{f_Y}(|\widehat{\vartheta}_k - \vartheta(f)|^2) \leq \|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|^2_{\mathbb{L}^1_{\mathbb{R}}} + \mathfrak{C}_g\frac{\sigma_Y}{n}\Delta_{\Psi,g}(k). \tag{5.6}$$

The first summand $\|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|^2_{\mathbb{L}^1(\mathbb{R})}$ is an upper bound of the squared bias term $|\mathbb{E}^n_{f_Y}(\widehat{\vartheta}_k) - \vartheta(f)|^2$ of our estimator. It is worth stressing out, that the squared bias term is vanishing for $k\to\infty$ but is not in general monotone decreasing. For the upcoming analysis of the risk, the bound $\|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|^2_{\mathbb{L}^1(\mathbb{R})}$ simplifies the considerations since its both, monotone decreasing and convergences to 0. Exploiting (5.5) the following consistency result can be deduced whose proof is omitted.

Corollary 5.2.10 (Consistency):
In addition to (5.3), assume that $\sigma_Y := \mathbb{E}_{f_Y}(Y_1^{2(c-1)}) < \infty$. Then for any sequence $(k_n)_{n\in\mathbb{N}}$ with $k_n\to\infty$ and $\|\mathbb{1}_{[-k_n,k_n]}\Psi/\mathcal{M}_c[g]\|^2_{\mathbb{L}^1(\mathbb{R})}n^{-1}\to 0$ for $n\to 0$ holds

$$\mathbb{E}^n_{f_Y}(|\widehat{\vartheta}_{k_n} - \vartheta(f)|^2) \to 0,$$

implying that $|\widehat{\vartheta}_{k_n} - \vartheta(f)|^2 \xrightarrow{\mathbb{P}} 0$.

Remark 5.2.11:
Despite the fact, that the first bound (5.5) only requires a finite second moment of X^{c-1} and U^{c-1} in many cases we have $\Delta_{\Psi,g}(k)\|\mathbb{1}_{[-k.k]}\Psi/\mathcal{M}_c[g]\|^{-2}_{\mathbb{L}^1(\mathbb{R})} \to 0$ as $k\to\infty$. In other words the bound on the variance term in (5.6) increases slower in k than the bound presented in (5.5). It is worth stressing out, that there exist cases where the opposite effect occurs. For instance let the error U be Lognormal distribution with parameter $\mu = 0, \lambda = 1$, see Example 2.1.4. Then $\sup_{u\in\mathbb{R}_+} u^{2c-1}g(u) = \sup_{z\in\mathbb{R}}(2\pi)^{-1/2}\exp(2(c-1)z)\exp(-z^2/2)) < \infty$. Thus if $\mathbb{E}_f(X^{2(c-1)}) < \infty$ both bounds are finite and following the argumentation of Butucea and Tsybakov (2008) we see, that in the special case of point-wise density estimation, the inequality presented in (5.5) is more favorable than the inequality presented in (5.6). □

For the upcoming theory we focus on the bound (5.6) of Proposition 5.2.9. Keep in mind that under $\|gx^{2c-1}\|_\infty < \infty$ the order of the second summand, also referred as variance term, is essentially determined by the order of $\Delta_{\Psi,g}(k)$ as k tends to infinity.

5.3 Minimax theory

Let us briefly revisit the example of the estimation of the survival function $S : \mathbb{R}_+ \to [0, 1]$ based on direct i.i.d. observations $(X_j)_{j\in[\![n]\!]}$, discussed in Section 4.1. Here, it is well-known that the empirical survival function,

$$\widehat{S}_X(x) = n^{-1}\sum_{j=1}^{n} \mathbb{1}_{(0,X_i)}(x), \quad x \in \mathbb{R}_+,$$

fulfills both

$$\mathbb{E}^n_f(|\widehat{S}_X - S\|^2_{x^0}) \leq C(f)n^{-1} \text{ and } \mathbb{E}^n_f(\|\widehat{S}_X(x_o) - S(x_o)|^2) \leq C(f, x_o)n^{-1}$$

under suitable moment assumptions. In other words, the empirical survival function attempts a parametric rate n^{-1} for both risks. In Section 4.1, we have seen that for the mean integrated squared error risk, there are examples of error distributions $g : \mathbb{R}_+ \to \mathbb{R}_+$ such that even in case of indirect observations $(Y_j)_{j\in[\![n]\!]}$, a parametric rate, respectively a almost parametric rate, is achievable with the proposed estimator $\widehat{S}_k$, compare Proposition 4.1.5. There we have seen, that parametric rates for the MISE can be deduced if $\mathcal{M}_c[g]$, the Mellin transform of g, converges not to fast to zero. A similar effect occurs when considering the MSE. More generally we will see, that for the estimation of $\vartheta(f)$ the interplay between the functions Ψ and $\mathcal{M}_c[g]$ determines whenever a parametric, or almost parametric, rate is achievable.

The parametric case In this paragraph we distinguish between two cases when due to Proposition 5.2.9 the rate of convergence of the estimator is parametric or eventually non-parametric. Note that the upper bound (5.6) of Proposition 5.2.9 is only informative, if $\Delta_{\Psi,g}(k) < \infty$ for all $k \in \mathbb{R}_+$ which we assume from now on.

(P) If $\sup_{k\in\mathbb{R}_+} \Delta_{\Psi,g}(k) = \|\Psi/\mathcal{M}_c[g]\|^2_{\mathbb{L}^2(\mathbb{R})} < \infty$, i.e. the second summand in (5.6) is uniformly bounded in k, then the rate is parametric. Precisely, since $\Psi\mathcal{M}_c[f] \in \mathbb{L}^1_{\mathbb{R}}$ for each $n \in \mathbb{N}$ there exists a cut-off parameter $k_n \in \mathbb{R}_+$ such that $\|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|^2_{\mathbb{L}^1_{\mathbb{R}}} \leq n^{-1}$ by dominated convergence. Obviously, due to (5.6) of Proposition 5.2.9 for all $k \in [k_n, \infty)$ we have

$$\mathbb{E}^n_{f_Y}(|\widehat{\vartheta}_k - \vartheta(f)|^2) \leq (1 + \|f\|_{1,x^{2(c-1)}}\|g\|_{\infty,x^{2c-1}}\|\Psi/\mathcal{M}_c[g]\|^2_{\mathbb{L}^2(\mathbb{R})})\, n^{-1}. \tag{5.7}$$

(NP) Let $\sup_{k\in\mathbb{R}_+} \Delta_{\Psi,g}(k) = \infty$, i.e. the second summand in (5.6) is unbounded as $k \to \infty$. If $\|\mathbb{1}_{(K,\infty)}\Psi\mathcal{M}_c[f]\|^2_{\mathbb{L}^1_{\mathbb{R}}} = 0$ for some $K \in \mathbb{R}_+$ then the rate is again parametric, i.e., we obtain a similar bound to (5.7) for $k = K$ and replacing $\|\Psi/\mathcal{M}_c[g]\|^2_{\mathbb{L}^2(\mathbb{R})}$ by $\Delta_{\Psi,g}(K)$. On the other hand if $\|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|^2_{\mathbb{L}^1_{\mathbb{R}}} > 0$ for all $k \in \mathbb{R}_+$, then Proposition 5.2.9 (5.6) necessitates an optimal choice $k_n := \arg\inf\{\|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|^2_{\mathbb{L}^1_{\mathbb{R}}} + \Delta_{\Psi,g}(k)n^{-1} : k \in \mathbb{R}_+\}$ of the cut-off parameter realising a squared-bias-variance trade-off. Note that $n\|\mathbb{1}_{(k_n,\infty)}\Psi\mathcal{M}_c[f]\|^2_{\mathbb{L}^1_{\mathbb{R}}} + \Delta_{\Psi,g}(k_n) \to \infty$ as $n \to \infty$ and hence the rate is non-parametric. □

Let us now again study the families of SMOOTH ERROR DENSITIES and SUPER SMOOTH ERROR DENSITIES introduced in Definition 3.2.4 and Definition 3.3.1.
The error density g is called SMOOTH if there exists a constant $C_g, c_g, \gamma \in \mathbb{R}_+$ such that

$$c_g(1+t^2)^{-\gamma/2} \leq \big|\mathcal{M}_c[g](t)\big| \leq C_g(1+t^2)^{-\gamma/2}, \quad t \in \mathbb{R} \tag{[G1]}$$

and it is referred to as SUPER SMOOTH if there exist constants $C_g, c_g, \lambda, \rho \in \mathbb{R}_+$ and $\gamma \in \mathbb{R}$ such that

$$c_g(1+t^2)^{-\gamma/2}\exp(-\lambda|t|^\rho) \leq \big|\mathcal{M}_c[g](t)\big| \leq C_g(1+t^2)^{-\gamma/2}\exp(-\lambda|t|^\rho),\ t \in \mathbb{R}. \tag{[G2]}$$

On the other hand to calculate the growth of $\Delta_{\Psi,g}$ we specify the decay of Ψ. Similar to the error density g we consider the case of a SMOOTH Ψ, i.e. there exists a constant $c_\Psi, C_\Psi, p \in \mathbb{R}$ such that

$$c_\Psi(1+t^2)^{-p/2} \leq \big|\Psi(t)\big| \leq C_\Psi(1+t^2)^{-p/2}, \quad t \in \mathbb{R}, \tag{[$\mathbf{\Psi1}$]}$$

and a SUPER SMOOTH Ψ, i.e. there exists $c_\Psi, C_\Psi, \mu, R \in \mathbb{R}_+$ and $p \in \mathbb{R}$ such that

$$c_\Psi(1+t^2)^{-p/2}\exp(-\mu|t|^R) \leq \big|\mathcal{M}_c[\Psi](t)\big| \leq C_\Psi(1+t^2)^{-p/2}\exp(-\mu|t|^R),\ t \in \mathbb{R}. \tag{[$\mathbf{\Psi2}$]}$$

As we see in the following illustration the examples of Ψ considered in Illustration Examples 5.2.1-5.2.6 do fit into these two cases.

Example 5.3.1 (Examples 5.2.1-5.2.6 continued)**:**
Let $x_o \in \mathbb{R}_+$.

(i) POINT-WISE DENSITY ESTIMATION:
[$\mathbf{\Psi1}$] is satisfied with $p = 0$, since $|\Psi(t)| = x_o^{-c}$.

(ii) POINT-WISE C.D.F. ESTIMATION :
[$\mathbf{\Psi1}$] is satisfied with $p = 1$ since $|\Psi(t)| = x_o^{1-c}|(1-c)^2 + t^2|^{-1/2}$.

(iii) POINT-WISE SURVIVAL FUNCTION ESTIMATION:
[$\mathbf{\Psi1}$] is satisfied with $p = 1$ since $|\Psi(t)| = x_o^{1-c}|(1-c)^2 + t^2|^{-1/2}$.

(iv) MEAN RESIDUAL LIFE ESTIMATION:
[$\mathbf{\Psi1}$] is satisfied with $p = 2$, since $|\Psi(t)| = x_o^{2-c}|((2-c)^2 + t^2)((1-c)^2 + t^2)|^{1/2}$.

(v) POINT-WISE LAPLACE TRANSFORM ESTIMATION.
[$\mathbf{\Psi2}$] is satisfied with $p = 1-2c$, $\mu = \pi/2$ and $R = 1$, since $|\Psi(t)| = t_o^{c-1}|\Gamma(1-c+it)|$.

(vi) POINT-WISE DERIVATIVE ESTIMATION.
[$\mathbf{\Psi1}$] is satisfied with $p = -1$, since $|\Psi(t)| = x_o^{-c-1}|c^2 + t^2|^{1/2}$.

Let us briefly summarize specifications for the parameters in [$\mathbf{\Psi1}$], [$\mathbf{\Psi2}$], [**G1**] and [**G2**] leading to the scenario (P), i.e. a parametric rate of convergence. The next assertion is an immediate consequence of Proposition 5.2.9 (5.6) (compare also the discussion of (P) above) and we omit the proof.

Corollary 5.3.2 (The parametric case):
In addition to (5.3), assume that $\mathbb{E}_{f_Y}(Y_1^{2(c-1)}) < \infty$ and $\|gx^{2c-1}\|_\infty < \infty$. Then scenario (P) occurs, i.e. the rate is parametric, in each of the three cases

(i) [$\mathbf{\Psi 1}$] and [**G1**], if $2p - 2\gamma > 1$;

(ii) [$\mathbf{\Psi 2}$] and [**G1**] or

(iii) [$\mathbf{\Psi 2}$] and [**G2**], if $(R > \rho)$, $(R = \rho, \mu > \lambda)$ or $(R = \rho, \mu = \lambda, 2p - 2\gamma > 1)$.

Let us briefly revisit the five examples for Ψ introduced in the Examples 5.2.1-5.2.6. Clearly, we get a parametric rate of convergence for point-wise estimation of the survival function and cumulative distribution function if the error density fulfills [**G1**] with $\gamma < 1/2$. In contrast the rate is parametric for point-wise estimation of the Laplace transform if the error density g fulfills either [**G1**] with $\gamma > 0$ or [**G2**] with $(\rho < 1)$, $(\rho = 1, \lambda < \pi/2)$, or $(\rho = 1, \lambda = \pi/2, \gamma < -c)$. For the mean residual lifetime a parametric rate is attainable if g satisfies [**G1**] with $\gamma < 3/2$.

The non-parametric case In the following section we develop the minimax theory for the plug-in spectral cut-off estimator under the assumptions [**G1**] and [$\mathbf{\Psi 1}$]. First we derive from Proposition 5.2.9 (5.6) a maximal upper bound uniformly over Mellin-Sobolev ellipsoids for linear functionals satisfying assumption [$\mathbf{\Psi 1}$]. Our main result of this section states separately a lower bound for each of the Examples 5.2.1, 5.2.3, 5.2.4 and 5.2.6, that is point-wise estimation of the density f, the survival function S, the the mean residual lifetime and the derivative f'. Let us restrict ourselves to the scenario [**G1**] and [$\mathbf{\Psi 1}$] with $2p - 2\gamma < 1$, i.e. a non-parametric rate of convergence (see Corollary 5.3.2). Clearly, in this situation we have $\Delta_{\Psi,g}(k) \leq C(g, \Psi)k^{2\gamma-2p+1}$ for all $k \in \mathbb{R}_+$. It remains to specify the order of the bias term. To do so, we consider Mellin-Sobolev ellipsoids at the development point $c \in \mathbb{R}$, defined in Proposition 3.1.4, given by

$$\mathbb{W}_c^s(L) := \{h \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) : |h|_{s,c}^2 := \|(1+t^2)^{s/2}\mathcal{M}_c[h]\|_{\mathbb{R}}^2 < L\}. \tag{5.8}$$

We denote the associated subset of densities by

$$\mathbb{D}_c^s(L) := \{f \in \mathbb{W}_c^s(L) : f \text{ is a density}, \mathbb{E}_f(X_1^{2(c-1)}) \leq L\}. \tag{5.9}$$

Similar smoothness classes have been considered by Butucea and Comte (2009), Mabon (2016), Pensky (2017) and Kappus (2014) for the estimation of linear functionals. Although the examples in Examples 5.2.1-5.2.6 may suggest to consider local smoothness classes, our general estimation strategy allows us to consider also linear functional other than point evaluation. Now, for each $f \in \mathbb{D}_c^s(L)$ and Ψ satisfying [$\mathbf{\Psi 1}$] we clearly have

$$\|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|_{\mathbb{L}^1(\mathbb{R})}^2 \leq L\int_{(k,\infty)} |\Psi(t)|^2(c^2+t^2)^{-s}dt \leq C(L,\Psi,c)k^{-2s-2p+1}. \tag{5.10}$$

Consequently, the next result is a direct consequence of Proposition 5.2.9 (5.6) and we omit its proof.

Theorem 5.3.3 (Upper bound for the minimax risk)**:**
In addition to (5.3) and $\mathbb{E}_g(U_1^{2(c-1)}) < \infty$, $\|gx^{2c-1}\|_\infty < \infty$ assume that g and Ψ satisfy [**G1**] and [$\mathbf{\Psi 1}$], respectively, with $2p - 2\gamma < 1$. If $s > 1/2 - p$ selecting $k_n := n^{1/(2s+2\gamma)}$ as cut-off parameter we obtain

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_{f_Y}((\widehat{\vartheta}_{k_n} - \vartheta(f))^2) \leq C(L, \Psi, g)\, n^{-(2s+2p-1)/(2s+2\gamma)}.$$

Note that $s > 1/2 - p$ implies $\Psi\mathcal{M}_c[f] \in \mathbb{L}^1(\mathbb{R})$ by a simple calculus given in the proof of Theorem 5.3.3 in the Section 5.6. Before stating matching lower bounds let us briefly illustrate the last result using the Examples 5.2.1-5.2.4 and 5.2.6.

Example 5.3.4 (Examples 5.2.1-5.2.4 and 5.2.6 continued)**:**
Let $x_o \in \mathbb{R}$.

(i) Point-wise density estimation:
Since $p = 0$ for each $s > 1/2 = 1/2 - p$ from Theorem 5.3.3 follows

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_{f_Y}((\widehat{\vartheta}_{k_n} - \vartheta(f))^2) \leq x_o^{-2c} C(L, \Psi, g)\, n^{-(2s-1)/(2s+2\gamma)}.$$

(ii) Point-wise c.d.f. estimation:
Since $p = 1$ for each $s \geq 0 > 1/2 - p$ and $\gamma \geq 1/2$ from Theorem 5.3.3 follows (with $c < 1$)

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_{f_Y}((\widehat{\vartheta}_{k_n} - \vartheta(f))^2) \leq x_o^{2-2c} C(L, \Psi, g)\, n^{-(2s+1)/(2s+2\gamma)}.$$

Note that for $\gamma < 1/2$ the rate is parametric due to Corollary 5.3.2 (see also (P)).

(iii) Point-wise survival function estimation:
Since $p = 1$ for each $s \geq 0 > 1/2 - p$ and $\gamma \geq 1/2$ from Theorem 5.3.3 follows (with $c > 1$)

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_{f_Y}((\widehat{\vartheta}_{k_n} - \vartheta(f))^2) \leq x_o^{2-2c} C(L, \Psi, g)\, n^{-(2s+1)/(2s+2\gamma)}.$$

Note that for $\gamma < 1/2$ the rate is parametric due to Corollary 5.3.2 (see also (P)).

(iv) Mean residual life estimation:
Since $p = 2$ for each $s \geq 0 > 1/2 - p$ and $\gamma \geq 3/2$ from Theorem 5.3.3 follows (with $c > 2$)

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_{f_Y}((\widehat{\vartheta}_{k_n} - \vartheta(f))^2) \leq x_o^{4-2c} C(L, \Psi, g)\, n^{-(2s+3)/(2s+2\gamma)}.$$

(v) Point-wise derivative estimation:
Since $p = -1$ for each $s > 3/2 = 1/2 - p$ from Theorem 5.3.3 follows (with $c \in \mathbb{R}$)

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_{f_Y}((\widehat{\vartheta}_{k_n} - \vartheta(f))^2) \leq x_o^{-2-2c} C(L, \Psi, g)\, n^{-(2s-3)/(2s+2\gamma)}.$$

We shall emphasize that in Example 5.3.4 (i) the sign of c changes drastically the impact of the evaluation point on the upper bound. In fact, for $c > 0$ it appears that the estimation in a point

x_o close to 0 is harder than for larger values of x_o. In contrast, if $c < 0$ the effect is just inverted. Further in Example 5.3.4 (iii) the estimator seems to have a better behavior far away from zero. An opposite behavior can be found for the estimation of the c.d.f., compare Example 5.3.4 (ii). However, in Remark 5.2.7 we already indicated how to use an estimator of the survival function for the construction of an c.d.f. estimator and vice versa. The results of Example 5.3.4 suggests to estimate the survival function directly or using the c.d.f. estimator, if $x_o \in \mathbb{R}_+$ is close to 0 or not. For the estimation of the derivative, Example 5.3.4 (v), analogously to (i) the impact of the evaluation point changes for $c < 1$ and $c > 1$. (iv) confirms our expectation (compare with Example 5.2.4) that faster convergence rates than for point-wise estimation of the survival function in Example 5.3.4 (iii) are attainable.

Remark 5.3.5:
Interestingly, Belomestny and Goldenshluger (2020) derive for point-wise density estimation a rate of $n^{-2s/(2s+2\gamma+1)}$ under similar assumptions on the error density g. However, they consider a maximal risk over Hölder-type regularity classes rather than Mellin-Sobolev spaces. Roughly speaking, Hölder-type regularity classes characterize regularity rather locally compared to the global nature of Mellin-Sobolev spaces. Even if the rates in Illustration 5.3.4 may seem less sharp compared to Belomestny and Goldenshluger (2020), they cannot be improved over Mellin-Sobolev spaces as shown by the lower bounds below. □

Lower bound For the following part, we impose additional assumption on the error density g. In fact, we suppose that g has bounded support, i.e. there is $d \in \mathbb{R}_+$ such that $g(x) = 0$ for all $x \in (d, \infty)$. For the sake of simplicity suppose $d = 1$. Furthermore let the Mellin transform $\mathcal{M}_{1/2}[g]$ of g at the development point $c = 1/2$ satisfy

$$c_g(1+t^2)^{-\gamma/2} \leq |\mathcal{M}_{1/2}[g](t)| \leq C_g(1+t^2)^{-\gamma/2}, \quad t \in \mathbb{R}. \qquad (\textbf{[G1']})$$

For technical reasons we restrict ourselves to the case of $c > 1/2$.

Theorem 5.3.6 (Lower bound for the minimax risk)**:**
Let $s, \gamma \in \mathbb{N}$ and assume g satisfies [**G1**] and [**G1'**].

(i) Point-wise density estimation: Then there exist constants $c_{g,x_o,1}, L_{s,g,x_o,c,1} > 0$, such that for all $L \geq L_{s,g,x_o,c,1}$, $n \in \mathbb{N}$ and for any estimator $\widehat{\vartheta}(f)$ of $\vartheta(f) = f(x_o)$ based on an i.i.d. sample $Y_1, \dots, Y_n$,

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_{f_Y}^n((\widehat{\vartheta}(f) - \vartheta(f))^2) \geq c_{g,x_o,1} n^{-(2s-1)/(2s+2\gamma)}.$$

(ii) Point-wise survival function estimation: Then for $c > 1$ there exist constants $c_{g,x_o,2}, L_{s,g,x_o,c,2} > 0$, such that for all $L \geq L_{s,g,x_o,c,2}$, $n \in \mathbb{N}$ and for any estimator $\widehat{\vartheta}(f)$ of $\vartheta(f) = S(x_o)$ based on an i.i.d. sample $Y_1, \dots, Y_n$,

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_{f_Y}^n((\widehat{\vartheta}(f) - \vartheta(f))^2) \geq c_{g,x_o,2} n^{-(2s+1)/(2s+2\gamma)}.$$

(iii) Point-wise derivative estimation: Then there exist constants $c_{g,x_o,3}, L_{s,g,x_o,c,3} > 0$, such that for all $L \geq L_{s,g,x_o,c,3}$, $n \in \mathbb{N}$ and for any estimator$\widehat{\vartheta}(f)$ of $\vartheta(f) = f'(x_o)$ based on an i.i.d. sample $Y_1, \dots, Y_n$,

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_{f_Y}^n((\widehat{\vartheta}(f) - \vartheta(f))^2) \geq c_{g,x_o,3} n^{-(2s-3)/(2s+2\gamma)}.$$

(iv) Mean residual lifetime: Then for $c > 2$ there exist constants $c_{g,x_o,4}, L_{s,g,x_o,c,4} > 0$, such that for all $L \geq L_{s,g,x_o,c,4}$, $n \in \mathbb{N}$ and for any estimator $\widehat{\vartheta}(f)$ of $\vartheta(f) = \mathbb{E}_f((X - x_o)\mathbb{1}_{X \geq x_o})$ based on an i.i.d. sample $Y_1, \dots, Y_n$,

$$\sup_{f \in \mathbb{D}_c^s(L)} \mathbb{E}_{f_Y}^n((\widehat{\vartheta}(f) - \vartheta(f))^2) \geq c_{g,x_o,4} n^{-(2s+3)/(2s+2\gamma)}.$$

Let us highlight that in multiplicative censoring the family of $\mathrm{Beta}_{(1,k)}$ error densities fulfills both assumption [**G1**] and [**G1'**].

Faster rates Revisiting Section 3.1.3 we get that for the Examples 5.2.1-5.2.6, the family of Gamma and Log-Normal distributions satisfies

$$(\vartheta_k - \vartheta(f))^2 \leq \left| \int_{(k,\infty)} |\Psi(t)| \mathcal{M}_c[f](t) d\lambda(t) \right|^2 \leq C_{f,\Psi} \exp(-\alpha k^a)$$

for $\alpha \in \mathbb{R}_+$, dependent on f and $a = 1$ for the Gamma distribution and $a = 2$ for the Log-Normal distribution and $C_{f,\Psi}$ a finite positive constant. Now choosing $k_n = \log(n)^{1/a}/\alpha^{-1}$ we can ensures that in the case of $2p - 2\gamma < 1$ we still get

$$\mathbb{E}_{f_Y}^n((\widehat{\vartheta}_{k_n} - \vartheta(f))^2) \leq C_{f,\Psi,g} \frac{\log(n)^{(2\gamma-2p+1)/a}}{n}$$

leading to a faster rate than provided by Theorem 5.3.3 for any choice of $s > 1/2 - p \in \mathbb{R}_+$. In the next section, we therefore present a data-driven method in order to choose the parameter $k \in \mathbb{R}_+$ based only on the sample $(Y_j)_{j \in [\![n]\!]}$.

5.4 Data-driven method

We now focus on the case (NP), that is $\sup_{k\in\mathbb{R}_+}\Delta_{\Psi,g}(k)=\infty$. In this scenario the first summand in the upper bound (5.6) of Proposition 5.2.9 is decreasing in k while the second summand is increasing and unbounded. A choice of the parameter $k\in\mathbb{R}_+$ realizing an optimal trade-off is thus non-trivial. We therefore define a data-driven procedure for the choice of the parameter $k\in\mathbb{R}_+$ inspired by the work of Goldenshluger and Lepski (2011). In fact, let us reduce the set of eligible dimension parameters to $[\![K_n]\!]$ with

$$K_n := \max\big\{k\in[\![n^{1/2}(\log(n))^{-3}]\!] : \frac{\Delta_{\Psi,g}(k)}{n}\leq \Delta_{\Psi,g}(1)\big\}. \tag{5.11}$$

We further introduce the variance term up to a $(\log n)$-term

$$V(k) := \chi\mathfrak{C}_g\sigma_Y\Delta_{\Psi,g}(k)(\log n)n^{-1} \tag{5.12}$$

with $\sigma_Y=\mathbb{E}_{f_Y}(Y_1^{2(c-1)})$ and $\mathfrak{C}_g=\|gx^{2c-1}\|_\infty/\mathbb{E}_g(U_1^{2c-1})$ defined in Proposition 5.2.9 and $\chi>0$, a positive constant chosen later on. Comparing estimators with different cut-off parameters we estimate the bias term by

$$A(k) := \max_{k'\in[\![k+1,K_n]\!]}((\widehat{\vartheta}_{k'}-\widehat{\vartheta}_k)^2-V(k'))_+. \tag{5.13}$$

Since the term σ_Y in $V(k)$ depending on the density f is unknown, we replace it by its empirical counterpart $\widehat{\sigma}_Y := n^{-1}\sum_{j\in[\![n]\!]}Y_j^{2(c-1)}$. Summarizing we estimate $V(k)$ and $A(k)$ by

$$\widehat{V}(k) := \frac{2\chi\mathfrak{C}_g}{n}\widehat{\sigma}_Y\Delta_{\Psi,g}(k)\log(n) \text{ and } \widehat{A}(k) := \max_{k'\in[\![k+1,K_n]\!]}((\widehat{\vartheta}_{k'}-\widehat{\vartheta}_k)^2-\widehat{V}(k'))_+. \tag{5.14}$$

Below we study the estimator $\widehat{\vartheta}_{\widehat{k}}$ of $\vartheta(f)$ with cut-off parameter selected by

$$\widehat{k} := \arg\min_{k\in[\![K_n]\!]}(\widehat{A}(k)+\widehat{V}(k)) \tag{5.15}$$

which depends only on observable quantities and hence $\widehat{\vartheta}_{\widehat{k}}$ is completely data-driven.

Theorem 5.4.1 (Data-driven choice of $k\in\mathbb{R}_+$)**:**
In addition to (5.3), suppose that $\mathbb{E}_{f_Y}(Y_1^{8(c-1)})<\infty$ and $\|gx^{2c-1}\|_\infty<\infty$. Then for any $\chi\geq 72$ there is a positive constant $C(f,g,\psi,\chi)$ such that

$$\mathbb{E}^n_{f_Y}((\vartheta(f)-\widehat{\vartheta}_{\widehat{k}})^2)\leq 9\inf_{k\in[\![K_n]\!]}\Big(\|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|^2_{\mathbb{L}^1_{\mathbb{R}}}+V(k)\Big)+C(f,g,\psi,\chi)\,n^{-1}.$$

Let us briefly discuss the last result. Consider the plug-in spectral cut-off estimator $\widehat{\vartheta}_{k_n}$ with optimally chosen cut-off parameter $k_n := \arg\inf\{\|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|^2_{\mathbb{L}^1_{\mathbb{R}}}+\Delta_{\Psi,g}(k)n^{-1} : k\in\mathbb{R}_+\}$. Its risk is up to a constant bounded by $\inf\{\|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|^2_{\mathbb{L}^1_{\mathbb{R}}}+\Delta_{\Psi,g}(k)n^{-1} : k\in\mathbb{R}_+\}$ and we show below for its maximal risk over certain Mellin-Sobolev balls a matching lower bound. However, the oracle inequality given in Theorem 5.4.1 essentially provides an upper bound for the risk of the fully data-driven estimator up to a constant by $\min\{\|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|^2_{\mathbb{L}^1_{\mathbb{R}}}+\Delta_{\Psi,g}(k)(\log n)n^{-1}$:

$k \in [\![K_n]\!]\}$. Typically we have $k_n \leq K_n$ for sufficiently large sample sizes $n \in \mathbb{N}$ such that both upper bounds essentially differ only by the additional $(\log n)$-factor. Its appearance is a known fact in the context of local estimation and it is widely considered as an acceptable price for adaptation (in the context of non-parametric Gaussian regression it is unavoidable as shown in Brown and Low (1996)). Let us further comment on the additional moment assumptions of Theorem 5.4.1. For $c \in \mathbb{R}$ close to one, the apparently high moment assumption $\mathbb{E}_{f_Y}(Y_1^{8(c-1)}) < \infty$ is rather weak. For point-wise density estimation, compare Example 5.2.1, this assumption is trivially satisfied if $c = 1$ is selected. For the point-wise estimation of the survival function (respectively, cumulative distribution function), it is not possible to select $c = 1$ but values of $c \in \mathbb{R}$ arbitrary close to one are possible.
If $\gamma \geq 1$ then the optimal cut-off parameter in Theorem 5.3.3 satisfies $k_n = \lfloor n^{1/(2s+2\gamma)} \rfloor \leq k^{1/2}$ and thus $k_n \in [\![K_n]\!]$ as introduced in (5.11) . The next result follows immediately from Theorem 5.4.1 and (5.10), and we omit the proof.

Corollary 5.4.2:
In addition to (5.3), $\mathbb{E}_{f_Y}(Y_1^{8(c-1)}) < \infty$ and $\|gx^{2c-1}\|_\infty < \infty$ assume that g and Ψ satisfy [**G1**] and [$\boldsymbol{\Psi}$**1**], respectively, with $2p - 2\gamma < 1$. If $s > 1/2 - p$ then for each $f \in \mathbb{D}_c^s(L)$ we have

$$\mathbb{E}_{f_Y}(|\widehat{\vartheta}_{\widehat{k}} - \vartheta(f)|^2) \leq C(f, g, \Psi)(\log(n)/n)^{(2s+2p-1)/(2s+2\gamma)}.$$

5.5 Conclusion

In this section, we have considered the local estimation of a variety of quantities of interest, for example density, survival function and derivative of the density, under multiplicative measurement error. Here, the term local estimation refers to the fact that we considered a point wise risk in contrary to the global risk studied in Chapter 3 and Chapter 4. More precisely, we have seen that these local estimations can be interpreted as examples of the estimation of the evaluation $\vartheta(f)$ of a linear functional ϑ at the density f. We then proposed an estimator based on a plug-in estimation, that is $\widehat{\vartheta}_k := \vartheta(\widehat{f}_k)$, where $\widehat{f}_k$ is the density estimator studied in Chapter 3. The theoretical properties of our proposed estimator are derived in terms of minimax-optimality of the estimation strategy, compare Theorem 5.3.3 and Theorem 5.3.6 and the study of a fully data-driven estimator $\widehat{\vartheta}_{\widehat{k}}$, see Theorem 5.4.1.
Considering several examples for the linear functional, we were able to compare different quantities of interest in terms of well-definedness and non-parametric rates. Throughout the chapter, we focus manly on the case of smooth error densities and excluded for the sake of simplicity oscillating and super smooth error densities from our study, which has been analyzed in Section 3.4 and 3.3. Furthermore, a linear functional model for multivariate densities or for stationary samples, like in Section 3.5 and 3.6 are left open for further research. Another compelling research direction would be to study the case of an unknown error distribution g of the measurement error U.

5.6 Proofs

The following Bernstein inequality, which in this form can be found in Comte (2017), is based on a similar formulation in Birgé and Massart (1998).

Lemma 5.6.1 (Bernstein inequality):
Let $(T_j)_{j\in[\![n]\!]}$ be independent random variables with $|T_j| \leq b$ and $\mathbb{V}\mathrm{ar}(T_j) \leq v^2$ for all $j \in [\![n]\!]$. Then for all $\eta \in \mathbb{R}_+$ we have

$$\mathbb{P}\Big(\big|\sum_{j\in[\![n]\!]}(T_j - \mathbb{E}(T_j))\big| \geq n\eta\Big) \leq 2\exp\Big(-\min\big(\tfrac{n\eta^2}{4v^2}, \tfrac{n\eta}{4b}\big)\Big).$$

Proof of Proposition 5.2.9. For $k \in \mathbb{R}_+$ denoting $\vartheta_k := \mathbb{E}^n_{f_Y}(\widehat{\vartheta}_k)$ the usual squared bias-variance decomposition gives

$$\mathbb{E}^n_{f_Y}((\widehat{\vartheta}_k - \vartheta(f))^2) = (\vartheta_k - \vartheta(f))^2 + \mathbb{V}\mathrm{ar}^n_{f_Y}(\widehat{\vartheta}_k) \tag{5.16}$$

where we bound each right hand side term separately. By an application of the Fubini-Tonelli theorem we have $\vartheta_k = (2\pi)^{-1}\int_{[-k,k]}\Psi(-t)\mathcal{M}_c[f](t)d\lambda(t)$ and the first summand in (5.16) is bounded by

$$(\vartheta_k - \vartheta(f))^2 = \left(\frac{1}{2\pi}\int_{[-k,k]^c}\Psi(-t)\mathcal{M}_c[f](t)d\lambda(t)\right)^2 \leq \|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|^2_{\mathbb{L}^1(\mathbb{R})}.$$

To estimate the second summand in (5.16), we note that

$$4\pi^2 n\mathbb{V}\mathrm{ar}^n_{f_Y}(\widehat{\vartheta}_k) \leq \mathbb{E}_{f_Y}\left(\left|\int_{[-k,k]}\frac{\Psi(-t)Y_1^{c-1+it}}{\mathcal{M}_c[g](t)}d\lambda(t)\right|^2\right) \leq \mathbb{E}_{f_Y}\left(\left|\int_{[-k,k]}\frac{|\Psi(t)|Y_1^{c-1}}{|\mathcal{M}_c[g](t)|}d\lambda(t)\right|^2\right).$$

Combining (5.16) with the last bounds we deduce the assertions (5.5) and (5.6) of the proposition. (5.5) follows immediately from the identities

$$\mathbb{E}_{f_Y}\left(\left|\int_{[-k,k]}\frac{|\Psi(t)|Y_1^{c-1}}{|\mathcal{M}_c[g](t)|}d\lambda(t)\right|^2\right) = \mathbb{E}_{f_Y}(Y_1^{2(c-1)})\|\mathbb{1}_{[-k,k]}\Psi/\mathcal{M}_c[g]\|^2_{\mathbb{L}^1(\mathbb{R})}$$

and $\mathbb{E}_{f_Y}(Y_1^{2(c-1)}) = \sigma_Y$. Consider (5.6). Exploiting the Plancherel identity (2.3.5) we have

$$\begin{aligned}
\mathbb{E}_{f_Y}\left(\left|\frac{1}{2\pi}\int_{[-k,k]}\frac{\Psi(-t)Y_1^{c-1+it}}{\mathcal{M}_c[g](t)}d\lambda(t)\right|^2\right) &= \int_{\mathbb{R}_+} f_Y(y)y^{2c-1}\left|\mathcal{M}^\dagger_{1/2}\left[\frac{\Psi\mathbb{1}_{[-k,k]}}{\mathcal{M}_c[g]}\right](y)\right|^2 d\lambda(y)\\
&\leq \|f_Y x^{2c-1}\|_\infty \left\|\mathcal{M}^\dagger_{1/2}\left[\frac{\Psi\mathbb{1}_{[-k,k]}}{\mathcal{M}_c[g]}\right]\right\|^2_{x^0}\\
&= \frac{\|f_Y x^{2c-1}\|_\infty}{2\pi}\|\mathbb{1}_{[-k,k]}\Psi/\mathcal{M}_c[g]\|^2_{\mathbb{L}^2(\mathbb{R})}\\
&\leq \frac{\mathfrak{C}_g\sigma_Y}{2\pi}\|\mathbb{1}_{[-k,k]}\Psi/\mathcal{M}_c[g]\|^2_{\mathbb{L}^2_{\mathbb{R}}}
\end{aligned} \tag{5.17}$$

which implies (5.6). Indeed, we have $\|f_Y\|_{\infty,x^{2c-1}} \leq \mathfrak{C}_g\sigma_Y$, since for any $y > 0$

$$y^{2c-1}f_Y(y) = \int_{\mathbb{R}_+} f(x)x^{2c-1}g(y/x)\frac{y^{2c-1}}{x^{2c-1}}dx \leq \|gx^{2c-1}\|_\infty\mathbb{E}_f(X_1^{2(c-1)}) = \mathfrak{C}_g\mathbb{E}_{f_Y}(Y_1^{2(c-1)})$$

and the proof is complete. □

Proof of Theorem 5.3.6. First we outline here the main steps of the proof. We construct two densities f_o, f_1 in $\mathbb{D}_c^s(L)$ by a perturbation with a small bump, such that the difference $(\vartheta(f_1) - \vartheta(f_o))^2$ and the Kullback-Leibler divergence of their induced distributions can be bounded from below and above, respectively. The claim follows then by applying Theorem 2.2 in Tsybakov (2009). We use the following construction,
We set $f_o(x) := \exp(-x)$ for $x \in \mathbb{R}_+$. Let $C_o^\infty(\mathbb{R})$ be the set of all infinitely differentiable functions with compact support in $\mathbb{R}$ and let $\varphi \in C_o^\infty(\mathbb{R})$ with

$$\operatorname{supp}[\varphi] \subset [-1,1], \quad \int_{[-1,1]} \varphi(x) d\lambda(x) = 0 \quad \text{and} \quad C_{\infty,\varphi}(j) := \max(\|\varphi^{(\ell)}\|_\infty, \ell \in [\![0,j]\!]).$$

Let us further consider the operator $\mathrm{S} : C_o^\infty(\mathbb{R}) \to C_o^\infty(\mathbb{R})$ with $\mathrm{S}[f](x) = -xf^{(1)}(x)$ for all $x \in \mathbb{R}$ and define $\mathrm{S}^1 := \mathrm{S}$ and $\mathrm{S}^n := \mathrm{S} \circ \mathrm{S}^{n-1}$ for $n \in \mathbb{N}, n \geq 2$. For each $x_o \in \mathbb{R}_+$, $h \in (0, \min(x_o/2, 1))$ (to be selected below) and $j \in \mathbb{N}$ we define the bump-function

$$\varphi_{j,h,x_o} := \mathrm{S}^j[\varphi_{h,x_o}], \text{ where } \varphi_{h,x_o}(x) := \varphi(\frac{x_o - x}{h}), \ x \in \mathbb{R}.$$

A direct calculus shows that $\varphi_{j,h,x_o}(x) = \sum_{i=1}^{j} c_{i,j} x^i h^{-i} \varphi^{(i)}(\frac{x_o - x}{h})$ for $x \in \mathbb{R}_+$ and $c_{i,j} \geq 1$. For a bump-amplitude $\delta > 0$ and $\gamma \in \mathbb{N}$ we define

$$f_1(x) = f_o(x) + \delta h^{\gamma+s-1/2} \varphi_{\gamma,h,x_o}(x), \quad x \in \mathbb{R}_+. \tag{5.18}$$

The next lemma implies that f_1 is a density for δ chosen sufficiently small.

Lemma 5.6.2:
For any $0 < \delta < \delta_o(\varphi, \gamma, x_o) := \exp(-3x_o/2)(3x_o/2)^{-\gamma}(C_{\varphi,\infty}(\gamma)c_\gamma)^{-1}$ the function f_1, defined in (3.34), is a density, where $c_\gamma = \sum_{i=1}^{\gamma} c_{i,\gamma}$.

Further we show that f_o and f_1 lie inside the ellipsoids $\mathbb{D}_c^s(L)$ for L big enough. This is captured in the following lemma.

Lemma 5.6.3:
Let $s \in \mathbb{N}$ and $c > 1/2$. Then, for all $L \geq L_{s,c,\gamma,\delta,\varphi,x_o} > 0$ holds f_o and f_1, as in (3.34), belong to $\mathbb{D}_c^s(L)$.

For sake of simplicity we denote for a function $h \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ the multiplicative convolution with g by $\widetilde{h} := [h * g]$.

Lemma 5.6.4:
$\mathrm{KL}(\widetilde{f}_1, \widetilde{f}_0) \leq C_{g,x_o,\varphi,f_o} \delta^2 h^{2s+2\gamma}$ where KL is the Kullback-Leibler divergence.

Selecting $h = n^{-1/(2s+2\gamma)}$, it follows

$$\frac{1}{M} \sum_{j=1}^{M} \mathrm{KL}(\widetilde{f}_1^{\otimes n}, \widetilde{f}_o^{\otimes n}) = \frac{n}{M} \sum_{j=1}^{M} \mathrm{KL}(\widetilde{f}_1, \widetilde{f}_o) \leq \delta^2 C_{g,x_o,\varphi,f_o} < 1/8$$

for all $\delta \leq \delta_{g,x_o,\varphi,f_o}$.

Lemma 5.6.5:
Let us additionally assume that $\varphi^{(\gamma-p)}(0)>0$, where $p=0$ (point-wise density estimation), $p=1$ (point-wise survival function estimation), $p=2$ (mean residual life) and $p=-1$ (point wise derivative estimation). Then in all four cases, there exists $h_{\varphi,x_o,\gamma}>0$ such that for all $h\leq h_{\gamma,x_o,\varphi}$ holds

$$(\vartheta(f_1)-\vartheta(f_o))^2\geq c_{\varphi,\gamma,x_o,\delta}h^{2s+2p-1}.$$

Thereby, we can use Theorem 2.2 of Tsybakov (2009), which in turn for any estimator $\widehat{\vartheta}(f)$ of $\vartheta(f)$ implies

$$\sup_{f\in\mathbb{D}_c^s(L)}\mathbb{P}^n_{f_Y}\left((\widehat{\vartheta}(f)-\vartheta(f))^2\geq\frac{c_{\varphi,\delta,x_o,\gamma}}{2}n^{-(2s+2p-1)/(2s+2\gamma)}\right)>\frac{3}{8}.$$

Note that the constant $c_{\varphi,\delta,x_o,\gamma}$ does only depend on φ,δ,x_o and γ, hence it is independent of the parameters s,L and n. The claim of Theorem 5.3.6 follows by using Markov's inequality, which completes the proof. □

Proof of Lemma 5.6.2. For any $h\in C_o^\infty(\mathbb{R})$ holds $\mathrm{S}[h]\in C_o^\infty(\mathbb{R})$ and thus $\mathrm{S}^j[h]\in C_o^\infty(\mathbb{R})$ for any $j\in\mathbb{N}$. Further for $h\in C_o^\infty(\mathbb{R}_+)$ holds $\int_{\mathbb{R}}\mathrm{S}[h](x)d\lambda(x)=\int_{\mathbb{R}}h(x)d\lambda(x)$ which implies that for any $\delta>0$ and we have $\int_{\mathbb{R}_+}f_1(x)d\lambda(x)=1$.
By construction (5.18) $\mathrm{supp}(\varphi_{h,x_o})$ in $[x_o/2,3x_o/2]$. Since $\mathrm{supp}(\mathrm{S}[h])\subseteq\mathrm{supp}(h)$ for all $h\in C_o^\infty(\mathbb{R})$ the function φ_{γ,h,x_o} has support in $[x_o/2,3x_o/2]$ too. First, for $x\notin[x_o/2,3x_o/2]$ holds $f_1(x)=\exp(-x)\geq 0$. Further for $x\in[x_o/2,3x_o/2]$ holds

$$f_1(x)=f_o(x)+\delta h^{s+\gamma-1/2}\varphi_{\gamma,h,x_o}(x)\geq\exp(-3x_o/2)-\delta(3x_o/2)^\gamma C_{\varphi,\infty}(\gamma)c_\gamma$$

since $\|\varphi_{j,h,x_o}\|_\infty\leq(3x_o/2)^jC_{\varphi,\infty}(j)c_jh^{-j}$ for any $s\geq1$ and $j\in\mathbb{N}$ where $c_j:=\sum_{i=1}^jc_{i,j}$. Now choosing $\delta\leq\delta_o(\varphi,\gamma,x_o):=\exp(-3x_o/2)(3x_o/2)^{-\gamma}(C_{\varphi,\infty}(\gamma)c_\gamma)^{-1}$ ensures $f_1(x)\geq0$ for all $x\in\mathbb{R}_+$. □

Proof of Lemma 5.6.3. Our proof starts with the observation that for all $t\in\mathbb{R}$ and $c>0$ it holds that

$$\mathcal{M}_c[f_o](t)\sim t^{c-1/2}\exp(-|t|\pi/2),\quad |t|\geq2,$$

by applying the Stirling formula, compare Belomestny and Goldenshluger (2020). Thus for every $s\in\mathbb{N}$ there exists $L_{s,c}$ such that $|f_o|^2_{s,c}\leq L$ for all $L\geq L_{s,c}$.
Next we consider $|f_o-f_1|_{s,c}$. We have $|f_o-f_1|^2_s=\delta^2h^{2s+2\gamma-1}|\varphi_{\gamma,h,x_o}|^2_{s,c}$ where $|\,.\,|_{s,c}$ is defined in (5.8). Now since $\mathrm{supp}(\varphi_{\gamma,h,x_o})\subset[x_o/2,3x_o/2]$ and $\varphi_{\gamma,h,x_o}\in C_o^\infty(\mathbb{R})$ we have that its Mellin transform is well-defined for any $c\in\mathbb{R}$. Then Proposition 2.3.12 implies $|\mathcal{M}_c[\varphi_{\gamma+s,h,x_o}](t)|^2=(c^2+t^2)^s|\mathcal{M}_c[\varphi_{\gamma,h,x_o}](t)|^2$ and thus

$$|\varphi_{\gamma,h,x_o}|^2_{s,c}\leq C_c\int_{\mathbb{R}}|\mathcal{M}_c[\varphi_{\gamma+s,h,x_o}](t)|^2d\lambda(t)=C_c\int_{[x_o/2,3x_o/2]}x^{2c-1}|\varphi_{\gamma+s,h,x_o}(x)|^2d\lambda(x)$$

by the Parseval identity, Proposition 2.3.5, which implies that $|\varphi_{\gamma,h,x_o}|^2_{s,c}\leq C_{c,x_o}\|\varphi_{\gamma+s,h,x_o}\|^2_{\mathbb{R}}$. Now applying the Jensen's inequality leads to

$$\|\varphi_{\gamma+s,h,x_o}\|^2_{\mathbb{R}}\leq C_{\gamma,s}\sum_{j=1}^{\gamma+s}h^{-2j}\int_{[x_o-h,x_o+h]}x^{2j}\varphi^{(j)}(\frac{x-x_o}{h})^2d\lambda(x)\leq C_{\gamma,s,x_o}h^{-2\gamma-2s+1}C_{\varphi,\infty}(\gamma+s).$$

Thus $|f_o - f_1|^2_{s,c} \leq C_{(c,s,\gamma,\delta,\varphi,x_o)}$ and $|f_1|^2_{s,c} \leq 2(|f_o - f_1|^2_{s,c} + |f_1|^2_{s,c}) \leq 2(C_{(c,s,\gamma,\delta,\varphi,x_o)} + L_{s,c}) =: L_{s,c,\gamma,\delta,\varphi,x_o,1}$. Now let us consider the moment condition. First we see that $\int_0^\infty x^{2(c-1)} f_o(x) = C_c$. Further since $h < x_o/2$ we have

$$\delta h^{s+\gamma-1/2} \int_{\mathbb{R}_+} x^{2(c-1)} \varphi_{\gamma,h,x_o}(x) d\lambda(x) \leq C_{\gamma,\delta} \sum_{j=1}^{\gamma} h^{s+\gamma+1/2-j} \int_{[x_o/2,3x_o/2]} x^{2(c-1)+j} \varphi^{(j)}(\frac{x-x_o}{h}) d\lambda(x)$$
$$\leq C_{s,c,\gamma,\delta,\varphi,x_o}$$

Thus we have $\mathbb{E}_{f_o}(X^{2(c-1)}), \mathbb{E}_{f_1}(X^{2(c-1)}) \leq C_c + C_{s,c,\gamma,\delta,\varphi,x_o} =: L_{s,c,\gamma,\delta,\varphi,x_o,2}$ Choosing now $L_{s,c,\gamma,\delta,\varphi,x_o} = \max(L_{s,c,\gamma,\delta,\varphi,x_o,1}, L_{s,c,\gamma,\delta,\varphi,x_o,2})$ shows the claim. □

Proof of Lemma 5.6.4.
Since $\mathrm{KL}(\widetilde{f}_1, \widetilde{f}_o) \leq \chi^2(\widetilde{f}_1, \widetilde{f}_o) := \int_{\mathbb{R}_+} (\widetilde{f}_1(x) - \widetilde{f}_o(x))^2/\widetilde{f}_o(x) d\lambda(x)$ it is sufficient to bound the χ-squared divergence. We notice that $\widetilde{f}_{\boldsymbol{\theta}} - \widetilde{f}_o$ has support in $[0, 3x_o/2]$ since $f_1 - f_o$ has support in $[x_o/2, 3x_o/2]$ and g has support in $[0,1]$. In fact for $x > 3x_o/2$ holds $\widetilde{f}_{\boldsymbol{\theta}}(y) - \widetilde{f}_o(y) = \int_y^\infty (f_{\boldsymbol{\theta}} - f_o)(x) x^{-1} g(y/x) d\lambda(x) = 0$. Since f_o is monotone decreasing we can deduce that $\widetilde{f}_o$ is montone decreasing since for $x_2 \geq x_1 \in \mathbb{R}_+$ holds

$$\widetilde{f}_o(x_2) = \int_{(0,1)} g(x) x^{-1} f_o(x_2/x) d\lambda(x) \leq \int_{(0,1)} g(x) x^{-1} f_o(x_1/x) d\lambda(x) = \widetilde{f}_o(x_1)$$

since the integrand is strictly positive. We conclude therefore that there exists a constant $c_{f_o,x_o,g} > 0$ such that $\widetilde{f}_o(x) \geq c_{f_o,x_o,g} > 0$ for all $x \in (0, 3x_o/2)$. Thus

$$\chi^2(\widetilde{f}_1, \widetilde{f}_o) \leq \widetilde{f}_o(3x_o/2)^{-1} \|\widetilde{f}_1 - \widetilde{f}_o\|^2_{\mathbb{R}} = \widetilde{f}_o(3x_o/2)^{-1} \delta^2 h^{2s+2\gamma-1} \|\widetilde{\varphi}_{\gamma,h,x_o}\|^2_{\mathbb{R}}.$$

Let us now consider $\|\widetilde{\varphi}_{\gamma,h,x_o}\|^2_{\mathbb{R}}$. In the first step we see by application of the Plancherel identity, Proposition 2.3.5, that $\|\widetilde{\varphi}_{\gamma,h,x_o}\|^2_{\mathbb{R}} = \frac{1}{2\pi} \int_{\mathbb{R}} |\mathcal{M}_{1/2}[\widetilde{\varphi}_{\gamma,h,x_o}](t)|^2 d\lambda(t)$. Now for $t \in \mathbb{R}$, we see by using the multiplication theorem for Mellin transforms that $\mathcal{M}_{1/2}[\widetilde{\varphi}_{\gamma,h,x_o}](t) = \mathcal{M}_{1/2}[g](t) \cdot \mathcal{M}_{1/2}[\mathrm{S}^\gamma[\varphi_{h,x_o}]](t)$. Again we have $\mathcal{M}_{1/2}[\mathrm{S}^\gamma[\varphi_{h,x_o}]](t) = (1/2 + it)^\gamma \mathcal{M}_{1/2}[\varphi_{h,x_o}](t)$. Together with assumption **[G1']** we get

$$\|\widetilde{\varphi}_{\gamma,h,x_o}\|^2_{\mathbb{R}} \leq \frac{C_g}{2\pi} \int_{-\infty}^{\infty} |\mathcal{M}_{1/2}[\varphi_{h,x_o}](t)|^2 dt = C_g \|\varphi_{h,x_o}\|^2_{\mathbb{R}} \leq C_{g,x_o} h \|\varphi\|^2_{\mathbb{R}}.$$

Since $M \geq 2^K$ we conclude $\mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}^{(j)}}, \widetilde{f}_{\boldsymbol{\theta}^{(0)}}) \leq C_{g,x_o,\varphi,f_o} \delta^2 h^{2s+2\gamma}$. □

Proof of Lemma 5.6.5. In all cases, we have $(\vartheta(f_1) - \vartheta(f_o))^2 = \delta^2 h^{2\gamma+2s-1} (\vartheta(\varphi_{\gamma,h,x_o}))^2$. Since $\varphi_{\gamma,h,x_o}(x) = \sum_{i=1}^{\gamma} c_{i,\gamma} h^{-i} x^i \varphi^{(i)}(\frac{x_o - x}{h})$ we deduce

$$(\vartheta(\varphi_{\gamma,h,x_o}))^2 = \sum_{j,i=1}^{\gamma} c_{i,\gamma} c_{j,\gamma} h^{-i-j} \vartheta\left(x^j \varphi^{(j)}(\frac{x_o - x}{h})\right) \vartheta\left(x^i \varphi^{(i)}(\frac{x_o - x}{h})\right)$$
$$=: \Sigma + c^2_{\gamma,\gamma} h^{-2\gamma} \left(\vartheta(x^\gamma \varphi^{(\gamma)}(\frac{x_o - x}{h}))\right)^2 \geq c^2_{\gamma,\gamma} h^{-2\gamma} \left(\vartheta(x^\gamma \varphi^{(\gamma)}(\frac{x_o - x}{h}))\right)^2 - |\Sigma|.$$

It is now sufficient to show that there exist $c_{\varphi,x_o,\gamma}, C_{\varphi,x_o,\gamma} > 0$ and $p \in \mathbb{N}$ such that for h small enough

$$\sup_{j \in [\![1,\gamma]\!]} |\vartheta(x^j \varphi^{(j)}(\frac{x_o - x}{h}))| \leq C_{\varphi,x_o,\gamma} h^p \quad \text{and } |\vartheta(x^\gamma \varphi^{(\gamma)}(\frac{x_o - x}{h}))| \geq c_{\varphi,x_o,\gamma} h^p. \tag{5.19}$$

Indeed, if (5.19) holds true, then there exists $h_{\varphi,\gamma,x_o} > 0$ such that for all $h < h_{\varphi,\gamma,x_o}$ holds $|\Sigma| \leq \frac{c_{\gamma,\gamma}^2}{2} h^{-2\gamma} \left(\vartheta(x^\gamma \varphi^{(\gamma)}(\frac{x_o - x}{h}))\right)^2$ implying that

$$(\vartheta(\varphi_{\gamma,h,x_o}))^2 \geq \frac{c_{\gamma,\gamma}^2}{2} h^{-2\gamma} \left(\vartheta(x^\gamma \varphi^{(\gamma)}(\frac{x_o - x}{h}))\right)^2 \geq c_{\varphi,x_o,\gamma} h^{-2\gamma+2p}.$$

This shows the claim. It remains to prove (5.19) for each different choice of Ψ.
$\underline{(i)}$: It holds for $j \in [\![\gamma]\!]$ that

$$|\vartheta(x^j \varphi^{(j)}(\frac{x_o - x}{h})| = x_o^j |\varphi^{(j)}(0)| \leq C_{\varphi,x_o,\gamma}\, h^0$$

and $|\vartheta(x^\gamma \varphi^{(\gamma)}(\frac{x_o - x}{h}))| = x_o^\gamma \varphi^{(\gamma)}(0) =: c_{\varphi,x_o,\gamma} h^0$ since $\varphi^{(\gamma)}(0) > 0$.
$\underline{(ii)}$: Since $\varphi^{(j)}(\frac{x_o - x}{h})$ has support in $[x_o - h, x_o + h]$, where $h \leq x_o/2$, we have for $j \in [\![1, \gamma]\!]$

$$\begin{aligned}\left|\int_{(x_o,\infty)} x^j \varphi^{(j)}(\frac{x_o - x}{h}) d\lambda(x)\right| &\leq \int_{(x_o,x_o+h)} x^j |\varphi^{(j)}(\frac{x_o - x}{h})| d\lambda(x) \\ &\leq \left(\frac{3x_o}{2}\right)^j h \|\varphi^{(j)}\|_\infty \leq C_{\varphi,x_o,\gamma} h.\end{aligned}$$

Further, we get by an integration by parts

$$\begin{aligned}\int_{(x_o,\infty)} x^\gamma \varphi^{(\gamma)}(\frac{x_o - x}{h}) d\lambda(x) &= h x_o^\gamma \varphi^{(\gamma-1)}(0) + h\gamma \int_{(x_o,\infty)} x^{\gamma-1} \varphi^{(\gamma-1)}(\frac{x_o - x}{h})) d\lambda(x) \\ &\geq h x_o^\gamma \varphi^{(\gamma-1)}(0) - h^2 C_{\varphi,x_o,\gamma} \geq \frac{h x_o^\gamma \varphi^{(\gamma-1)}(0)}{2} > 0\end{aligned}$$

for $h < h_{\varphi,x_o,\gamma}$ since $\varphi^{(\gamma-1)}(0) > 0$.
Analogously we can handle the case (iii) with $p = 2$ and (iv) with $p = -1$. □

Proof of Theorem 5.4.1. Setting $\vartheta := \vartheta(f)$ we observe that

$$(\vartheta - \widehat{\vartheta}_{\widehat{k}})^2 \leq 2(\vartheta - \widehat{\vartheta}_k)^2 + 2(\widehat{\vartheta}_k - \widehat{\vartheta}_{k\wedge\widehat{k}})^2 + 2(\widehat{\vartheta}_{k\wedge\widehat{k}} - \widehat{\vartheta}_{\widehat{k}})^2.$$

Exploiting successively the definition (5.14) of $\widehat{A}$, and secondly the definition (5.15) of $\widehat{k}$ we obtain

$$\begin{aligned}(\vartheta - \widehat{\vartheta}_{\widehat{k}})^2 &\leq 2(\vartheta - \widehat{\vartheta}_k)^2 + 2(\widehat{A}(\widehat{k}) + \widehat{V}(k) + \widehat{A}(k) + \widehat{V}(\widehat{k})) \\ &\leq 2(\vartheta - \widehat{\vartheta}_k)^2 + 4(\widehat{A}(k) + \widehat{V}(k)).\end{aligned}$$

It is easy to check that $\widehat{A}(k) \leq A(k) + \max\{(V(k') - \widehat{V}(k'))_+ : k' \in [\![1, K_n]\!]\}$, and hence

$$\begin{aligned}(\vartheta - \widehat{\vartheta}_{\widehat{k}})^2 &\leq 2(\vartheta - \widehat{\vartheta}_k)^2 + 4A(k) + 4(\widehat{V}(k) - 2V(k)) + 8V(k) + 4 \max_{k' \in [\![1,K_n]\!]} (V(k') - \widehat{V}(k'))_+ \\ &\leq 9V(k) + 4(\vartheta - \vartheta_k)^2 + 4(\widehat{V}(k) - 2V(k)) + 4 \max_{k' \in [\![1,K_n]\!]} (V(k') - \widehat{V}(k'))_+ \\ &+ 4((\vartheta_k - \widehat{\vartheta}_k)^2 - V(k)/6)_+ + 4A(k). \qquad (5.20)\end{aligned}$$

Further for $k' \in [\![k+1, K_n]\!]$ from $(\widehat{\vartheta}_{k'} - \widehat{\vartheta}_k)^2/3 \leq (\widehat{\vartheta}_{k'} - \vartheta_{k'})^2 + (\widehat{\vartheta}_k - \vartheta_k)^2 + (\vartheta_k - \vartheta_{k'})^2$ we conclude

$$A(k) \leq 3 \max_{k' \in [\![k+1,K_n]\!]} \left((\widehat{\vartheta}_k - \vartheta_{k'})^2 + (\widehat{\vartheta}_k - \vartheta_k)^2 - V(k')/3\right)_+ + 3 \max_{k' \in [\![k+1,K_n]\!]} (\vartheta_k - \vartheta_{k'})^2.$$

Since $V(k') \geq \frac{1}{2}V(k) + \frac{1}{2}V(k')$ for each $k' \in [\![k, K_n]\!]$ due to the monotonicity of $k \mapsto V(k)$ we deduce

$$A(k) \leq 6 \max_{k' \in [\![k,K_n]\!]} \Big((\widehat{\vartheta}_k - \vartheta_{k'})^2 - V(k')/6\Big)_+ + 3 \max_{k' \in [\![k+1,K_n]\!]} (\vartheta_k - \vartheta_{k'})^2. \tag{5.21}$$

The second summand in (5.21) is bounded for any $k' \in [\![k+1, K_n]\!]$ by

$$|\vartheta_k - \vartheta_{k'}| = \frac{1}{2\pi}\left|\int_{[-k',-k)\cup(k,k']} \Psi(-t)\mathcal{M}_c[f](t)d\lambda(t)\right| \leq \pi^{-1}\|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|_{\mathbb{R}}. \tag{5.22}$$

Combining (5.20) with (5.21), (5.22), $|\vartheta_k - \vartheta| \leq \pi^{-1}\|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|_{\mathbb{L}^1_{\mathbb{R}}}$ and $\mathbb{E}^n_{f_Y}(\widehat{V}(k)) = 2V(k)$ for each $k \in [\![1, K_n]\!]$ we obtain

$$\begin{aligned}\mathbb{E}^n_{f_Y}(\vartheta - \widehat{\vartheta}_{\widehat{k}})^2 \leq 2\|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|^2_{\mathbb{L}^1_{\mathbb{R}}} + 9V(k) + 4\mathbb{E}^n_{f_Y}\Big(\max_{k' \in [\![K_n]\!]}(V(k') - \widehat{V}(k'))_+\Big)\\ + 26\mathbb{E}^n_{f_Y}\Big(\max_{k' \in [\![K_n]\!]}\Big((\widehat{\vartheta}_k - \vartheta_{k'})^2 - V(k')/6\Big)_+\Big).\end{aligned} \tag{5.23}$$

We bound the last two summands in (5.23) with the help of Lemma 5.6.6 below.

Lemma 5.6.6:
Under the assumptions of Theorem 5.4.1 we have

(i) $\mathbb{E}^n_{f_Y}\Big(\max_{k \in [\![K_n]\!]}(V(k) - \widehat{V}(k))_+\Big) \leq C(\mathbb{E}_{f_Y}(Y_1^{5(c-1)}), \sigma_Y, \Psi, g)n^{-1}$;

(ii) $\mathbb{E}^n_{f_Y}\Big(\max_{k \in [\![K_n]\!]}\Big((\widehat{\vartheta}_k - \vartheta_k)^2 - V(k)/6\Big)_+\Big) \leq C(\mathbb{E}_{f_Y}(Y_1^{8(c-1)}), \|gx^{2c-1}\|_\infty, \sigma_Y, \Psi, g)n^{-1}$.

Consequently we have

$$\mathbb{E}^n_{f_Y}(\vartheta - \widehat{\vartheta}_{\widehat{k}})^2 \leq 9 \inf_{k \in [\![K_n]\!]}\Big(\|\mathbb{1}_{(k,\infty)}\Psi\mathcal{M}_c[f]\|^2_{\mathbb{L}^1_{\mathbb{R}}} + V(k)\Big) + C(f, g, \psi)n^{-1}$$

which gives the desired conclusion and completes the proof. □

Proof of Lemma 5.6.6. Consider (i). On the event $\{|\sigma_Y - \widehat{\sigma}_Y| \leq \sigma_Y/2\}$ we have clearly $\sigma_Y \leq 2\widehat{\sigma}_Y$, and thus $(\sigma_Y - 2\widehat{\sigma}_Y)_+ \leq 2|\widehat{\sigma}_Y - \sigma_Y|\mathbb{1}_{\{|\widehat{\sigma}_Y - \sigma_Y| > \sigma_Y/2\}}$. Combining the last bound with the definition (5.12) and (5.14) of V and $\widehat{V}$, respectively, for each $k \in [\![K_n]\!]$ we obtain

$$\begin{aligned}(V(k) - \widehat{V}(k))_+ &= \frac{\chi}{n\log(n)}\mathfrak{C}_g(\sigma_Y - 2\widehat{\sigma}_Y)_+\Delta_{\Psi,g}(k)\\ &\leq \chi\mathfrak{C}_g\Delta_{\Psi,g}(1)\log(n)|\widehat{\sigma} - \sigma|\mathbb{1}_{\{|\widehat{\sigma}_Y - \sigma_Y| > \sigma_Y/2\}}\\ &= C(g, \Psi, \chi)\log(n)|\widehat{\sigma} - \sigma|\mathbb{1}_{\{|\widehat{\sigma}_Y - \sigma_Y| > \sigma_Y/2\}}\end{aligned}$$

using $\Delta_{\Psi,g}(k) \leq \Delta_{\Psi,g}(1)n$ for all $k \in [\![K_n]\!]$ due to (5.11). Applying the Cauchy-Schwarz inequality

$$\begin{aligned}\mathbb{E}^n_{f_Y}(|\widehat{\sigma}_Y - \sigma_Y|\mathbb{1}_{\{|\widehat{\sigma}_Y - \sigma_Y| > \sigma_Y/2\}}) &\leq \mathbb{V}\mathrm{ar}^n_{f_Y}(\widehat{\sigma}_Y)^{1/2}\mathbb{P}^n_{f_Y}(|\widehat{\sigma}_Y - \sigma_Y| > \sigma_Y/2)^{1/2}\\ &\leq C(\mathbb{E}_{f_Y}(Y_1^{4(c-1)})))n^{-1/2}\mathbb{P}^n_{f_Y}(|\widehat{\sigma}_Y - \sigma_Y| > \sigma_Y/2)^{1/2}.\end{aligned}$$

Now using the Nagaev inequality, compare Nagaev (1979), we get for $p = 3$ that

$$\mathbb{P}^n_{f_Y}(|\widehat{\sigma}_Y - \sigma_Y| > \sigma_Y/2)^{1/2} \leq C(\mathbb{E}_{f_Y}(Y_1^{5(c-1)}), \sigma_Y)n^{-1}$$

implying that

$$\mathbb{E}^n_{f_Y}\left(\max_{k\in[\![K_n]\!]}(V(k) - \widehat{V}(k))_+\right) \leq C(\Psi, g, \sigma_Y, \mathbb{E}_{f_Y}(Y_1^{5(c-1)}), \chi)n^{-1}. \tag{5.24}$$

Our proof of (ii) starts with a decomposition of $\widehat{\vartheta}_k - \vartheta_k$ into a sum of two parts. To control the last term we split the centered arithmetic mean $\widehat{\vartheta}_k - \vartheta_k$ into two terms, applying at one term a Bernstein inequality, cf Lemma 5.6.1, and standard techniques on the other term. For a positive sequence $(c_n)_{n\in\mathbb{N}}$ and $t \in \mathbb{R}$ introduce

$$\widehat{\mathcal{M}}_c(t) = n^{-1}\sum_{j=1}^n (Y_j^{c-1+it}\mathbb{1}_{(0,c_n)}(Y_j^{c-1}) + Y_j^{c-1+it}\mathbb{1}_{(c_n,\infty)}(Y_j^{c-1})) =: \widehat{\mathcal{M}}_{c,1}(t) + \widehat{\mathcal{M}}_{c,2}(t).$$

Split the centered arithmetic mean $\widehat{\vartheta}_k - \vartheta_k = \nu_{k,1} + \nu_{k,2}$ where $\nu_{k,i} := \frac{1}{2\pi}\int_{-k}^k \frac{\Psi(-t)}{\mathcal{M}_c[g](t)}(\widehat{\mathcal{M}}_{c,i}(t) - \mathbb{E}_{f_Y}(\widehat{\mathcal{M}}_{c,i}(t)))dt$. Thus we have

$$\max_{k\in[\![K_n]\!]}\left((\widehat{\vartheta}_k - \vartheta_k)^2 - V(k)/6\right)_+ \leq 2\max_{k\in[\![\mathcal{K}_n]\!]}\left(\nu_{k,1}^2 - \frac{1}{12}V(k)\right)_+ + 2\max_{k\in[\![\mathcal{K}_n]\!]}\nu_{k,2}^2.$$

For the first summand, we see that

$$\begin{aligned}
\mathbb{E}^n_{f_Y}\left(\max_{k'\in[\![K_n]\!]}(\nu_{k',1}^2 - \frac{1}{12}V(k'))_+\right) &\leq \sum_{k\in\mathcal{K}_n}\mathbb{E}^n_{f_Y}\left((\nu_{k,1}^2 - \frac{1}{12}V(k))_+\right)\\
&\leq \sum_{k\in\mathcal{K}_n}\int_{\mathbb{R}_+}\mathbb{P}^n_{f_Y}\left((\nu_{k,1}^2 - \frac{1}{12}V(k))_+ \geq x\right)d\lambda(x)\\
&\leq \sum_{k\in\mathcal{K}_n}\int_{\mathbb{R}_+}\mathbb{P}^n_{f_Y}\left(|\nu_{k,1}| \geq \sqrt{\frac{V(k)}{12} + x}\right)d\lambda(x).
\end{aligned}$$

Now our aim is to apply the Bernstein inequality 5.6.1. To do so, defining for $y > 0$ the function $h_k(y) := \frac{1}{2\pi}\int_{[-k,k]}\frac{\Psi(-t)}{\mathcal{M}_c[g](t)}y^{it}d\lambda(t)$ leads to

$$\nu_{k,1} = \frac{1}{n}\sum_{j=1}^n Y_j^{c-1}\mathbb{1}_{(0,c_n)}(Y_j^{c-1})h_k(Y_j) - \mathbb{E}_{f_Y}(Y_1^{c-1}\mathbb{1}_{(0,c_n)}(Y_1^{c-1})h_k(Y_1))$$

where $|h_k(y)| \leq (2\pi)^{-1}\int_{[-k,k]}\left|\frac{\Psi(t)}{\mathcal{M}_c[g](t)}\right|d\lambda(t) \leq \sqrt{k\Delta_{\Psi,g}(k)}$. Thus $|Y_j^{c-1}\mathbb{1}_{(0,c_n)}(Y_j^{c-1})h_k(Y_j)| \leq c_n\sqrt{k\Delta_{\Psi,g}(k)} =: b$. Further,

$$\mathbb{V}\mathrm{ar}_{f_Y}(Y_1^{c-1}\mathbb{1}_{(0,c_n)}(Y_1^{c-1})h_k(Y_1)) \leq \mathbb{E}_{f_Y}(Y_1^{2c-2}h_k^2(Y_1)) \leq \mathfrak{C}_g\sigma_Y\Delta_{\Psi,g}(k) =: v.$$

Therefore the Bernstein inequality yields, for any $x > 0$

$$\mathbb{P}^n_{f_Y}\left(|\nu_{k,1}| \geq \sqrt{\frac{V(k)}{12} + x}\right) \leq 2\max\left(\exp\left(-\frac{n}{4v}\left(\frac{V(k)}{12} + x\right)\right), \exp\left(-\frac{n}{8b}\left(\sqrt{\frac{V(k)}{12}} + \sqrt{x}\right)\right)\right)$$

using the concavity of the square root. We have thus to bound the 4 upcoming terms. In fact

$$\frac{n}{4v}\frac{V(k)}{12} = \frac{\chi}{48}\log(n) \geq \frac{3}{2}\log(n) \quad \text{and} \quad \frac{n}{4\nu} \geq \frac{C_{g,\Psi}}{4\sigma_Y}$$

for $\chi \geq 72$ which implies $\exp(-\frac{n}{4v}(\frac{V(k)}{12} + x)) \leq n^{-3/2}\exp(-xC_{g,\Psi}/4\sigma_Y)$. Moreover we have

$$\begin{aligned} \frac{n}{8b}\sqrt{\frac{V(k)}{12}} &= \frac{n\sqrt{\sigma_Y \mathfrak{C}_g \Delta_{\Psi,k}(k)\chi \log(n) n^{-1}}}{8c_n\sqrt{k\Delta_{\Psi,g}(k)12}} \\ &\geq \frac{\sqrt{n^{1/2}\sigma_Y \mathfrak{C}_g \log(n)}}{28c_n}\sqrt{\frac{n^{1/2}}{k}} \geq \frac{3}{2}\log(n) \end{aligned}$$

by definition of $\mathcal{K}_n$ and $c_n = \sqrt{n^{1/2}\sigma_Y \mathfrak{C}_g \log(n)}/42$. In analogy we can show that

$$\frac{n}{8b} = \frac{42n^{3/4}}{\sqrt{\sigma_Y \mathfrak{C}_g \log(n) k\Delta_{\Psi,g}(k)}} \geq \frac{5\sqrt{\log(n)}}{\sqrt{\sigma_Y C_{g,\Psi}}}$$

implying that $\exp(-\frac{n}{8b}(\sqrt{\frac{V(k)}{12}} + \sqrt{x})) \leq n^{-3/2}\exp(-5\sqrt{x\log(n)(\sigma_Y C_{g,\Psi})^{-1}})$. Thus we conclude

$$\begin{aligned} \mathbb{E}^n_{f_Y}\left(\max_{k' \in [\![K_n]\!]}(\nu^2_{k',1} - \frac{1}{12}V(k'))_+\right) &\leq \sum_{k \in [\![K_n]\!]} n^{-3/2}\int_{\mathbb{R}_+}\exp(-x\min(\frac{C_{g,\Psi}}{4\sigma}, 18\sqrt{\frac{\log(n)}{x\sigma_Y C_{g,\Psi}}}))d\lambda(x) \\ &\leq C(\sigma_Y, g, \Psi)\sum_{k \in [\![K_n]\!]} n^{-3/2} \leq \frac{C(\sigma_Y, g, \Psi)}{n}. \end{aligned}$$

For part *(ii)* we have $|\nu_{k',2}| \leq (2\pi)^{-1}\int_{[-k',k']}|\Psi(t)||\mathcal{M}_c[g](t)|^{-1}|\widehat{\mathcal{M}}_{c,2}(t) - \mathbb{E}_{f_Y}(\widehat{\mathcal{M}}_{c,2}(t))|d\lambda(t)$ implying with the Cauchy-Schwarz inequality that

$$\begin{aligned} \mathbb{E}^n_{f_Y}\left(\max_{k' \in [\![K_n]\!]}\nu^2_{k',2}\right) &\leq \mathbb{E}^n_{f_Y}((\frac{1}{2\pi}\int_{[-K_n,K_n]}\left|\frac{\Psi(t)}{\mathcal{M}_c[g](t)}\right||\widehat{\mathcal{M}}_{c,2}(t) - \mathbb{E}^n_{f_Y}(\widehat{\mathcal{M}}_{c,2}(t))|d\lambda(t))^2) \\ &\leq \frac{K_n}{2\pi}\Delta_{\Psi,g}(K_n)n^{-1}\mathbb{E}_{f_Y}(Y_1^{2c-2}\mathbb{1}_{(c_n,\infty)}(Y_1^{c-1})) \\ &\leq C_{\Psi,g}n^{1/2}\mathbb{E}_{f_Y}(Y_1^{(c-1)(2+u)})c_n^{-u} \end{aligned}$$

for any $u \in \mathbb{R}_+$. Choosing $u = 6$ leads to $\mathbb{E}_{f_Y}(\max_{k' \in [\![K_n]\!]}\nu^2_{k',2}) \leq C_{\Psi,g,\sigma}\mathbb{E}_{f_Y}(Y_1^{8(c-1)})n^{-1}$.

□

CHAPTER 6

Perspectives

In this work, we mainly consider the non parametric estimation of a quantity of interest of the distribution of a positive random variable in a multiplicative measurement errors model with known error densities. The following presented extension may be addressed in a future work.

Unknown error distribution In practice, the exact distribution of the error term is often unknown. There are several attempts to address this problem by estimating simultaneously the error distribution itself. In the literature, one consider for instance a parametric family of error distribution, a second sample of the error or panel data. In all of these models, additional knowledge (parametric family, additional samples) are necessary as the multiplicative convolution is commutative, that is

$$\mathbb{P}^X \odot \mathbb{P}^U = \mathbb{P}^Y = \mathbb{P}^U \odot \mathbb{P}^X.$$

In other words, without additional knowledge we cannot distinguish between signal and error.

Second sample In this scenario, we assume to have additionally to the sample $(Y_j)_{j\in[\![n]\!]}$ access to a second sample $(\widetilde{U}_j)_{j\in[\![m]\!]}$ from the error variable U. Based on this sample, we then derive an estimator $\widetilde{\mathcal{M}}_c[g]$ of the Mellin transform of U. Using this estimator a density estimator is given by

$$\widetilde{f}_k(x) := \frac{1}{2\pi}\int_{[-k,k]} x^{-c-it}\frac{\widehat{\mathcal{M}}_c(t)}{\widetilde{\mathcal{M}}_c[g](t)}dt, \quad x \in \mathbb{R}_+.$$

For this estimator, we can develop again minimax optimality and data-driven methods. Similar results can be found in Neumann and Hössjer (1997), Johannes (2009), Comte and Lacour (2010), Kappus and Mabon (2014) and Dattner et al. (2016).

Panel data 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 the same signal X m-times leading to observations $Y^1, \dots, Y^m$ which only differs in the measurement errors. For the sake of simplicity let us assume that $m = 2$. Inspired by Comte and Kappus (2015) and Kappus and Mabon (2014), we can deduce a representation of the Mellin transform of X in terms of the Mellin transform of $\boldsymbol{Y}$. Such a result for the Mellin transform can be used to deduce an estimator of the quantity of interest. Again, this estimator can be studied in terms of minimax optimality and data-driven choices of the smoothing parameters.

Non parametric testing Next to non parametric estimation, the problem of non parametric minimax and adaptive testing has been studied in the literature of statistical inverse problems. Given a null hypothesis density f_o, in non parametric testing we aim to study the testing problem

$$\mathrm{H}_o : f = f_o \quad \text{vs.} \quad \mathrm{H}_1 : f \in \mathcal{H}_1$$

given a sample of $Y = XU$ where $X \sim f$ and U is again a multiplicative measurement error.
In this scenario we observe that given a niveau $\alpha \in (0,1)$ and a power $\beta \in (0,1)$ the set $\mathcal{H}_1$ has to be of the form

$$\mathcal{H}_1 := \{f \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) : \|f - f_o\|^2_{x^{2c-1}} > r\}.$$

More precisely, we are in need of a gap between the null hypotheses H_o and H_1. Otherwise, each α-test will have zero power, that is for each α-test we are able to construct an alternative in such that the test has zero power. In non parametric minimax testing, we are now interested given$\alpha, \beta \in (0,1)$ how to large $r \in \mathbb{R}_+$ needs to be to allow the existence of a α-test with power β Based on an estimation of quadratic functional and the Mellin transform of f, we can then construct a test which is minimax optimal and propose a data-driven method for the choice of the upcoming smoothing parameter. This testing problem in the context of inverse statistical problems has been studied by Fromont and Laurent (2006), Butucea (2007), , Butucea et al. (2009), Schluttenhofer and Johannes (2020a), Schluttenhofer and Johannes (2020b),and Schluttenhofer and Johannes (2022).

Bibliography

Andersen, K. E. and Hansen, M. B. (2001). Multiplicative censoring: density estimation by a series expansion approach. *Journal of Statistical Planning and Inference*, 98(1-2):137–155.

Andrews, G. E., Askey, R., and Roy, R. (1999). *Special functions*, volume 71. Cambridge university press Cambridge.

Asgharian, M. and Wolfson, D. B. (2005). Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored prevalent cohort data. *The Annals of Statistics*, 33(5):2109–2131.

Asin, N. and Johannes, J. (2017). Adaptive nonparametric estimation in the presence of dependence. *Journal of Nonparametric Statistics*, 29(4):694–730.

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. 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. and Schoenmakers, J. (2016). Statistical inference for time-changed lévy processes via mellin transform approach. *Stochastic Processes and their Applications*, 126(7):2092–2122.

Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375.

Bradley, R. C. (1986). Basic properties of strong mixing conditions. In *Dependence in probability and statistics*, pages 165–192. Springer.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, 2:107–144.

Brenner Miguel, S. (2021). Multiplicative deconvolution estimator based on a ridge approach. *arXiv preprint arXiv:2108.01523*.

Brenner Miguel, S. (2022a). Anisotropic spectral cut-off estimation under multiplicative measurement errors. *Journal of Multivariate Analysis*, 190:104990.

Brenner Miguel, S. (2022b). Volatility density estimation by multiplicative deconvolution. *to be published: Journal of Nonparametric Statistics*.

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 errors. *Bernoulli*, 29(3):2247–2271.

Brenner Miguel, S. and Phandoidaen, N. (2022). Multiplicative deconvolution in survival analysis under dependency. *Statistics*, 56(2):297–328.

Brown, L. D. and Low, M. G. (1996). A constrained risk inequality with applications to nonparametric functional estimation. *The Annals of Statistics*, 24(6):2524–2535.

Brunel, E., Comte, F., and Genon-Catalot, V. (2016). Nonparametric density and survival function estimation in the multiplicative censoring model. *Test*, 25(3):570–590.

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., Matias, C., and Pouet, C. (2009). Adaptive goodness-of-fit testing from indirect observations. In *Annales de l'IHP Probabilités et statistiques*, volume 45, pages 352–372.

Butucea, C. and Tsybakov, A. B. (2008). Sharp optimality in density deconvolution with dominating bias. I. *Theory of Probability & Its Applications*, 52(1):24–39.

Comte, F. (2017). *Nonparametric estimation*. Master and Research. 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. and Genon-Catalot, V. (2006). Penalized projection estimator for volatility density. *Scandinavian Journal of Statistics*, 33(4):875–893.

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. (2010). Pointwise deconvolution with unknown error distribution. *Comptes rendus de l'Académie des sciences. Série I, Mathématique*, 348(5-6):323–326.

Comte, F. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, 49(2):569–609.

Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Canadian Journal of Statistics*, 34(3):431–452.

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407.

Dattner, I., Reiß, M., and Trabs, M. (2016). Adaptive quantile estimation in deconvolution with unknown error distribution. *Bernoulli*, 22(1):143 – 192.

Dedecker, J., Doukhan, P., Lang, G., José Rafael, L. R., Louhichi, S., and Prieur, C. (2007). Weak dependence. In *Weak dependence: With examples and applications*, pages 9–20. Springer.

Donoho, D. (1994). Statistical estimation and optimal recovery. *The Annals of Statistics*, 22:238–270.

Donoho, D. and Low, M. (1992). Renormalization exponents and optimal pointwise rates of convergence. *The Annals of Statistics*, 20:944–970.

Doukhan, P. (1994). *Mixing*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York. Properties and examples.

Durrett, R. (2019). *Probability: theory and examples*, volume 49. Cambridge university press.

Dussap, F. (2021). Anisotropic multivariate deconvolution using projection on the laguerre basis. *Journal of Statistical Planning and Inference*, 215:23–46.

Dussap, F. (2023). Nonparametric multiple regression by projection on non-compactly supported bases. *Annals of the Institute of Statistical Mathematics*, pages 1–41.

Efromovich, S. (1999). *Nonparametric curve estimation: methods, theory, and applications*. Springer Science & Business Media.

Engl, H. W., Hanke, M., and Neubauer, A. (1996). *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht.

Epstein, B. (1948). Some applications of the Mellin transform in statistics. *The Annals of Mathematical Statistics*, pages 370–379.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19(3):1257–1272.

Fox, C. (1957). Some applications of Mellin transforms to the theory of bivariate statistical distributions. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 53, pages 620–628. Cambridge University Press.

Fromont, M. and Laurent, B. (2006). Adaptive goodness-of-fit tests in a density model. *The Annals of Statistics*, 34(2):680 – 720.

Fryzlewicz, P. and Subba Rao, S. (2011). Mixing properties of ARCH and time-varying ARCH processes. *Bernoulli*, 17(1):320–346.

Gaver, D. P. and Lewis, P. A. W. (1980). First-order autoregressive gamma sequences and point processes. *Advances in Applied Probability*, 12(3):727–745.

Geenens, G. (2021). Mellin-Meijer kernel density estimation on $\mathbb{R}_+$. *Annals of the Institute of Statistical Mathematics*, 73(5):953–977.

Genon-Catalot, V., Jeantheau, T., and Laredo, C. (2003). Conditional likelihood estimators for hidden Markov models and stochastic volatility models. *Scandinavian Journal of Statistics*, 30(2):297–316.

Gloter, A. (2000). Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM: Probability and Statistics*, 4:205–227.

Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39:1608–1632.

Goldenshluger, A. and Pereverzev, S. V. (2000). Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probability Theory and Related Fields*, 118:169–186.

Guess, F. and Proschan, F. (1988). 12 mean residual life: theory and applications. *Handbook of statistics*, 7:215–224.

Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin*, pages 49–52.

Hall, P. and Meister, A. (2007). A ridge-parameter approach to deconvolution. *The Annals of Statistics*, 35(4):1535–1558.

Ibragimov, I. and Has'minskii, R. (1984). On nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory of Probability and its Applications*, 29:18–32.

Johannes, J. (2009). Deconvolution with unknown error distribution. *The Annals of Statistics*, 37(5A):2301 – 2323.

Kappus, J. (2014). Adaptive nonparametric estimation for Lévy processes observed at low frequency. *Stochastic Processes and their Applications*, 124(1):730–758.

Kappus, J. and Mabon, G. (2014). Adaptive density estimation in deconvolution problems with unknown error distribution. *Electronic Journal of Statistics*, 8(2):2879 – 2904.

Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *The Annals of Probability*, 33(3):1060–1077.

Laurent, B., Ludena, C., and Prieur, C. (2008). Adaptive estimation of linear functionals by model selection. *Electronic Journal of Statistics*, 2(993-1020).

Lepski, O. V. and Willer, T. (2019). Oracle inequalities and adaptive estimation in the convolution structure density model. *The Annals of Statistics*, 47(1):233 – 287.

Li, K. (1982). Minimaxity of the method of regularization of stochastic processes. *The Annals of Statistics*, 10:937–942.

Lindelöf, E. (1933). Robert Hjalmar Mellin. *Acta Mathematica*, 61(none).

Liu, W., Xiao, H., and Wu, W. B. (2013). Probability and moment inequalities under dependence. *Statistica Sinica*, pages 1257–1272.

Lomnicki, Z. (1967). On the distribution of products of random variables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 29(3):513–524.

Mabon, G. (2016). Adaptive deconvolution of linear functionals on the nonnegative real line. *Journal of Statistical Planning and Inference*, 178:1–23.

Meister, A. (2009). *Deconvolution problems in nonparametric statistics*, volume 193 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin.

Nagaev, S. V. (1979). Large deviations of sums of independent random variables. *The Annals of Probability*, pages 745–789.

Neumann, M. H. and Hössjer, O. (1997). On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7(4):307–330.

Pardoux, E. and Veretennikov, Y. (2001). On the Poisson equation and diffusion approximation. I. *The Annals of Probability*, 29(3):1061–1085.

Paris, R. B. and Kaminski, D. (2001). *Asymptotics and Mellin-Barnes integrals*, volume 85 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge.

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.

Rebelles, G. (2016). Structural adaptive deconvolution under $\lessdot_p$-losses. *Mathematical Methods of Statistics*, 25(1):26–53.

Rio, E. (2013). Inequalities and limit theorems for weakly dependent sequences.

Schluttenhofer, S. and Johannes, J. (2020a). Adaptive minimax testing for circular convolution. *Mathematical Methods of Statistics*, 29:106–133.

Schluttenhofer, S. and Johannes, J. (2020b). Adaptive minimax testing in inverse gaussian sequence space models. *arXiv preprint arXiv:2002.07623*.

Schluttenhofer, S. and Johannes, J. (2022). Adaptive pointwise density estimation under local differential privacy. *arXiv preprint arXiv:2206.07663*.

Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.

Speckman, P. (1979). Minimax estimation of linear functionals in a Hilbert space. Unpublished manuscript.

Springer, M. D. and Thompson, W. (1966). The distribution of products of independent random variables. *SIAM Journal on Applied Mathematics*, 14(3):511–526.

Springer, M. D. and Thompson, W. E. (1970). The distribution of products of beta, gamma and gaussian random variables. *SIAM Journal on Applied Mathematics*, 18(4):721–737.

Subrahmaniam, K. (1970). On some applications of Mellin transforms to statistics: Dependent random variables. *SIAM Journal on applied Mathematics*, 19(4):658–662.

Talagrand, M. (1996). New concentration inequalities in product spaces. *Inventiones mathematicae*, 126:505–563.

Triebel, H. (2006). *Theory of Function Spaces III*. Monographs in Mathematics. Birkhäuser Basel.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer Publishing Company, Incorporated.

van Es, B., Klaassen, C. A. J., and Oudshoorn, K. (2000). Survival analysis under cross-sectional sampling: length bias and multiplicative censoring. volume 91, pages 295–312. Prague Workshop on Perspectives in Modern Statistical Inference: Parametrics, Semi-parametrics, Non-parametrics (1998).

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.

Viennet, G. (1997). Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Related Fields*, 107(4):467–492.

Wells, W., Anderson, R., and Cell, J. W. (1962). The distribution of the product of two central or non-central chi-square variates. *The Annals of Mathematical Statistics*, pages 1016–1020.

Werner, D. (2006). *Funktionalanalysis*. Springer.

Wu, W. B. (2005). Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA*, 102(40):14150–14154.