Electronic Journal of Statistics
Vol. 18 (2024) 4795–4850
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2314

Multiplicative deconvolution under unknown error distribution

Sergio Brenner Miguel[1], Jan Johannes[2] and Maximilian Siebel[3]

[1] *Heidelberg University,*
e-mail: mail@sergiomiguel.de

[2] *Institute for Mathematics and Interdisciplinary Center for Scientific Computing, Heidelberg University,*
e-mail: johannes@math.uni-heidelberg.de

[3] *Institute for Mathematics, Heidelberg University,*
e-mail: siebel@math.uni-heidelberg.de

Abstract: We consider a multiplicative deconvolution problem, in which the density f or the survival function S^X of a strictly positive random variable X is estimated nonparametrically based on an i.i.d. sample from a noisy observation $Y = X \cdot U$ of X. The multiplicative measurement error U is supposed to be independent of X. The objective of this work is to construct a fully data-driven estimation procedure when the error density f^U is unknown. We assume that in addition to the i.i.d. sample from Y, we have at our disposal an additional i.i.d. sample drawn independently from the error distribution. The proposed estimation procedure combines the estimation of the Mellin transformation of the density f and a regularisation of the inverse of the Mellin transform by a spectral cut-off. The derived risk bounds and oracle-type inequalities cover both – the estimation of the density f as well as the survival function S^X. The main issue addressed in this work is the data-driven choice of the cut-off parameter using a model selection approach. We discuss conditions under which the fully data-driven estimator can attain the oracle-risk up to a constant without any previous knowledge of the error distribution. We compute convergences rates under classical smoothness assumptions. We illustrate the estimation strategy by a simulation study with different choices of distributions.

MSC2020 subject classifications: Primary 62G05; secondary 62G07, 62C20.
Keywords and phrases: Density and survival function estimation, multiplicative measurement errors, Mellin transform, Mellin-Sobolev space, inverse problem, adaptation.

Received August 2023.

Contents

arXiv: 2308.08423

1. Introduction

In this work let X and U be strictly positive and independent random variables both admitting unknown densities $f = f^X$ and f^U accordingly. We propose a data-driven estimation procedure for the density f as well as the survival function S^X of X based on an independent and identically distributed (i.i.d.) sample of size n from a multiplicative measurement model $Y = X \cdot U$ and an additional sample of size m drawn independently from the unknown error density f^U. In this situation Y admits also a density, given by

$$f^Y(y) := [f \circledast f^U](y) := \int_{\mathbb{R}_{>0}} f(x) f^U(y/x) x^{-1} \mathrm{d}x. \tag{1}$$

Estimating f from i.i.d. observations following the law of f^Y is a statistical inverse problem called multiplicative deconvolution. Multiplicative censoring is introduced and studied in [31] and [32]. It corresponds to the particular multiplicative deconvolution problem with multiplicative error U uniformly distributed on $[0, 1]$. [30] motivate and explain multiplicative censoring in survival analysis. [32] and [2] consider the estimation of the cumulative distribution function of X. Treating the model as an inverse problem [1] study series expansion methods. The density estimation in a multiplicative censoring model is considered in [13] using a kernel estimator and a convolution power kernel estimator. [15] analyse a projection density estimator with respect to the Laguerre basis assuming a uniform error distribution on an interval $[1-\alpha, 1+\alpha]$ for $\alpha \in (0, 1)$. A beta-distributed error U is studied in [4]. The multiplicative measurement error model covers all those three variations of multiplicative censoring. It was considered by [5] studying the point-wise density estimation and by [11] casting point-wise estimation as the estimation of the value of a known linear functional.

In contrast to local, global estimation of the density means estimating the density as a whole and measuring the accuracy of an estimator by an integrated mean squared error. Under multiplicative measurement errors it is considered in [10]. The authors use the Mellin transform and a spectral cut-off regularisation of its inverse to define an estimator for the unknown density f. In those three papers the key to the analysis of multiplicative deconvolution is the multiplication theorem, which for a density $f^Y = f \circledast f^U$ and their Mellin transformations $\mathrm{M}[f^Y]$, $\mathrm{M}[f]$ and $\mathrm{M}[f^U]$ (defined below) states $\mathrm{M}[f^Y] = \mathrm{M}[f]\,\mathrm{M}[f^U]$. It is used in [5] and [10] to construct respectively a kernel density estimator and a spectral cut-off estimator of the density f, while the later serves for a plug-in estimator in [11]. Under multiplicative measurement errors [8] studies the global multivariate density estimation while the global estimation of the survival function can be found in [12]. For local and global multiplicative deconvolution [5] and [10] both comment on the naive approach to apply standard additive deconvolution methods to the log-transformed data. Essentially, the additive deconvolution theorem for the log-transformed data equals a special case of the multiplicative convolution theorem. The authors point out that making use of the Mellin transform leads to a more flexible approach than the naive one.

We now turn to multiplicative deconvolution with unknown error density, which is inspired by similar ideas for additive deconvolution problems (see for instance [25] and [20]). Following the estimation strategy in [10] and [12], and borrowing ideas from the inverse problems community (see for instance [19]), we define spectral cut-off estimators $\hat{f}_k$ and $\widehat{S}_k^X$ of f and S^X, respectively, by replacing the unknown Mellin transformations of f^Y and f^U by empirical counterparts based on the two samples from U and Y and additional thresholding. The accuracy of the estimators $\hat{f}_k$ and $\widehat{S}_k^X$ are measured in terms of the global risk with respect to a weighted $\mathbb{L}^2$-norm on the positive real line $\mathbb{R}_{>0}$, i.e. a weighted integrated mean squared error. We observe that both global risks can be embedded into a more general risk analysis, which we then study in detail. The proposed estimation strategy depends on a further tuning parameter k, which has to be chosen by the user. In case of a known error density [10] and [12] propose a data-driven choice of the tuning parameter k by model selection exploiting the theory of [3], where we refer to [23] for an extensive overview. Our aim is to establish a fully data-driven estimation procedure for the density f and the survival function S^X when the error density is unknown and derive oracle-type upper risk bounds as well as convergences rates. A similar approach has been considered for additive deconvolution problems for instance in [16] and [21]. Regarding the two samples sizes n and m, by comparing the oracle-type risk bounds in the cases of known and unknown error densities, we characterise the influence of the estimation of the error density on the quality of the estimation. Interestingly, in case of additive convolution on the circle and the real line [21] and [16] derive respectively oracle-type inequalities with similar structure.

The paper is organised as follows. In Section 2 we start with recalling the definition of the Mellin transform as well stating certain properties. Secondly, supposing the error density to be known we review the spectral cut-off estimators of f and S^X as respectively proposed by [10] and [12]. We study their global risk with respect to weighted $\mathbb{L}^2$-norms and discuss how to generalise this risk analysis. Afterwards, we investigate oracle-type inequalities and minimax-optimal convergences rates under regularity assumptions on the Mellin transformations of f and f^U, respectively. In Sections 3 and 4 we dismiss the knowledge of the error density. In Section 3 we derive a general risk bound and oracle-type inequality. We construct in Section 4 the fully data-driven estimation procedure and eventually show a upper risk bound. The general theory applies in particular to the estimation of f and S^X. In Section 5 we illustrate the finite sample properties of the estimators proposed in Sections 3 and 4 by a simulation study with different choices of distributions.

2. Model assumptions and background

In the following paragraphs we introduce first the multiplicative measurement model with known error distribution and we recall the definition of the Mellin transform and its empirical counter part as well as their properties. Secondly, still assuming the error distribution is known in advance we briefly present a data-driven density estimation strategy due to [10] which we extend in the sequel to cover simultaneously the estimation of the density f as well as the survival function S^X of X.

2.1. The multiplicative measurement model

Let $(\mathbb{R}_{>0}, \mathscr{B}_{>0}, \lambda_{>0})$ denote the Lebesgue-measure space of all positive real numbers $\mathbb{R}_{>0}$ equipped with the restriction $\lambda_{>0}$ of the Lebesgue measure on the Borel σ-field $\mathscr{B}_{>0}$. Assume that X and U are independent random variables, both taking values in $\mathbb{R}_{>0}$ and admitting both a (Lebesgue) density $f := f^X$ and f^U, respectively. The multiplicative measurement model describes observations following the law of $Y := X \cdot U$. In this situation, Y admits a density $f^Y = f \circledast f^U$ given as multiplicative convolution of f and f^U, which can be computed explicitly as in (1). For a detailed discussion of multiplicative convolution and its properties as operator between function spaces we refer to [9]. However, assuming the error distribution is known, we have in the sequel access to an independent and identically distributed (i.i.d.) sample $\{Y_i\}_{i \in [\![n]\!]}$ drawn from f^Y, where we have used the shorthand notation $[\![a]\!] := [1, a] \cap \mathbb{N}$ for any $a \in \mathbb{R}_{\geq 1}$.

2.2. The (empirical) Mellin transform

In the subsequent, we keep the following objects and notations in mind: Given a density function v defined on $\mathbb{R}_{>0}$, i.e. a Borel-measurable nonnegative function $\mathrm{v} : \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$, let $\mathrm{v}\lambda_{>0}$ denote the σ-finite measure on $(\mathbb{R}_{>0}, \mathscr{B}_{>0})$ which is $\lambda_{>0}$ absolutely continuous and admits the Radon-Nikodym derivative v with respect to $\lambda_{>0}$. For $p \in [1,\infty]$ let $\mathbb{L}^p_+(\mathrm{v}) := \mathbb{L}^p(\mathbb{R}_{>0}, \mathscr{B}_{>0}, \mathrm{v}\lambda_{>0})$ denote the usual complex Banach-space of all (equivalence classes of) $\mathbb{L}^p$-integrable complex-valued function with respect to the measure $\mathrm{v}\lambda_{>0}$. Similarly, we set $\mathbb{L}^p(\mathrm{v}) := \mathbb{L}^p(\mathbb{R}, \mathscr{B}, \mathrm{v}\lambda)$ for a density function v defined on $\mathbb{R}$. If $\mathrm{v} = \mathbb{1}$, i.e. v is mapping constantly to one, we write shortly $\mathbb{L}^p_+ := \mathbb{L}^p_+(\mathbb{1})$ and $\mathbb{L}^p := \mathbb{L}^p(\mathbb{1})$. At this point we shall remark, that we have used and will further use the terminology *density*, whenever we are meaning a probability density function (such as f) and on the other hand side *density function*, whenever we are meaning a Radon-Nikodym derivative (such as v). For $c \in \mathbb{R}$ we introduce the density function $\mathrm{x}^c : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ given by $x \mapsto \mathrm{x}^c(x) := x^c$. The Mellin transform M_c is the unique linear and bounded operator between the Hilbert spaces $\mathbb{L}^2_+(\mathrm{x}^{2c-1})$ and $\mathbb{L}^2$, which for each $h \in \mathbb{L}^2_+(\mathrm{x}^{2c-1}) \cap \mathbb{L}^1_+(\mathrm{x}^{c-1})$ and $t \in \mathbb{R}$ satisfies

$$\mathrm{M}_c[h](t) := \int_{\mathbb{R}_{>0}} x^{c-1+\iota 2\pi t} h(x) \mathrm{d}\lambda_{>0}(x), \tag{2}$$

where $\iota \in \mathbb{C}$ denotes the imaginary unit. Similar to the Fourier transform the Mellin transform M_c is unitary, i.e. $\langle h, g \rangle_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})} = \langle \mathrm{M}_c[h], \mathrm{M}_c[g] \rangle_{\mathbb{L}^2}$ for any $h, g \in \mathbb{L}^2(\mathrm{x}^{2c-1})$. In particular, it satisfies a *Plancherel type identity*

$$\|h\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})} = \|\mathrm{M}_c[h]\|^2_{\mathbb{L}^2}, \quad \forall h \in \mathbb{L}^2_+(\mathrm{x}^{2c-1}). \tag{3}$$

Its adjoined (and inverse) $\mathrm{M}^\star_c : \mathbb{L}^2 \to \mathbb{L}^2_+(\mathrm{x}^{2c-1})$ fulfils for each $g \in \mathbb{L}^2 \cap \mathbb{L}^1$ and $x \in \mathbb{R}_{>0}$

$$\mathrm{M}^\star_c[g](x) := \int_{\mathbb{R}} x^{-c-\iota 2\pi t} g(t) \mathrm{d}\lambda(t). \tag{4}$$

For a detailed discussion of the Mellin transform and its properties we refer again to [9]. In analogy to the additive convolution theorem of the Fourier transform (see [24] for definitions and properties), there is a multiplicative convolution theorem for the Mellin transform. Namely, for any $h_1 \in \mathbb{L}^2_+(\mathrm{x}^{2c-1})$ and $h_2 \in \mathbb{L}^1_+(\mathrm{x}^{c-1}) \cap \mathbb{L}^2_+(\mathrm{x}^{2c-1})$ the multiplicative convolution theorem states

$$\mathrm{M}_c[h_1 \circledast h_2] = \mathrm{M}_c[h_1]\, \mathrm{M}_c[h_2]. \tag{5}$$

Here and subsequently, we assume that $f \in \mathbb{L}^2_+(\mathrm{x}^{2c-1})$ and $f^U \in \mathbb{L}^1_+(\mathrm{x}^{c-1}) \cap \mathbb{L}^2_+(\mathrm{x}^{2c-1})$, hence $f^Y \in \mathbb{L}^2_+(\mathrm{x}^{2c-1})$, for some $c \in \mathbb{R}$, which is from now on fixed. Note that $\mathbb{L}^2$-integrability is a common assumption in additive deconvolution, which might be here seen as the particular case $c = 1/2$. We should emphasise that the particular case $c = 0$ corresponds to the naive approach to log-transform the data and apply an additive deconvolution method. However, allowing for different values $c \in \mathbb{R}$ makes the dependence on $c \in \mathbb{R}$ of the assumptions stipulated below visible. For example for the naive approach, i.e. $c = 0$ the assumptions below translate to X^{-1} and U^{-1} having a finite second moment. Consequently, if the assumptions we impose below are satisfied for $c = 0$, then the results also hold true for the naive approach. In order to simplify the presentation in the following sections, we introduce some frequently used shorthand notations.

Notation 2.1. The Mellin transformations of the densities f^Y, f and f^U, respectively, we denote briefly by

$$\mathrm{M}_Y := \mathrm{M}_c[f^Y],\ \mathrm{M}_X := \mathrm{M}_c[f] \text{ and } \mathrm{M}_U := \mathrm{M}_c[f^U].$$

The reciprocal of a function $\mathrm{w} : \mathbb{R} \to \mathbb{C}$ is well-defined on the set $\{\mathrm{w} \neq 0\} := \{t \in \mathbb{R} : \mathrm{w}(t) \neq 0\}$ and for each $t \in \mathbb{R}$ we write for short

$$\mathrm{w}^\dagger(t) := \frac{1}{\mathrm{w}(t)} \mathbb{1}_{\{\mathrm{w}\neq 0\}}(t)$$

where $\mathbb{1}_{\{\mathrm{w}\neq 0\}}$ denotes an indicator function. Precisely, for any set $A \subseteq \mathbb{R}$ we write $\mathbb{1}_A$ for the indicator function, mapping $x \in \mathbb{R}$ to $\mathbb{1}_A(x) := 1$, if $x \in A$, and $\mathbb{1}_A(x) := 0$, otherwise. Since the Mellin transformation M_Y of f^Y is unknown we follow [10] and introduce an empirical counterpart based on the observations $\{Y_i\}_{i\in⟦n⟧}$. The *empirical Mellin transformation* $\widehat{\mathrm{M}}_Y$ is given by

$$\widehat{\mathrm{M}}_Y(t) := \frac{1}{n} \sum_{i\in⟦n⟧} Y_i^{c-1+\iota 2\pi t}$$

for any $t \in \mathbb{R}$. Observe that $\widehat{\mathrm{M}}_Y$ is a point-wise unbiased estimator of M_Y, i.e. for all $t \in \mathbb{R}$, we have $\mathbb{E}[\widehat{\mathrm{M}}_Y(t)] = \mathrm{M}_Y(t)$, where $\mathbb{E}$ denotes the expectation under the distribution of the observations $\{Y_i\}_{i\in⟦n⟧}$. Let us further introduce the point-wise scaled variance of $\widehat{\mathrm{M}}_Y$ defined for each $t \in \mathbb{R}$ by

$$\begin{aligned}\mathbb{V}_Y^2(t) := n\mathbb{E}[|\widehat{\mathrm{M}}_Y(t) - \mathrm{M}_Y(t)|^2] &= \mathbb{E}[|Y_1^{c-1+\iota 2\pi t} - \mathrm{M}_Y(t)|^2] \\ &\leq 1 + \mathbb{E}[Y_1^{2(c-1)}] =: \sigma_Y^2,\end{aligned}$$

whenever $\mathbb{E}[Y_1^{2(c-1)}]$ is finite. The estimation procedure we are discussing next is based on estimating the unknown Mellin transformation M_X in a first place. Having the multiplicative convolution theorem (5) in mind, a estimator of M_X is given by

$$\widetilde{\mathrm{M}}_X(t) := \widehat{\mathrm{M}}_Y(t)\, \mathrm{M}_U^\dagger(t) \tag{6}$$

for each $t \in \mathbb{R}$. To shorten notation we write $\widetilde{\mathrm{M}}_X^k := \widetilde{\mathrm{M}}_X \mathbb{1}_{[-k,k]}$ for each $k \in \mathbb{R}_{>0}$. Finally, in what follow we denote by $\mathfrak{C} \in \mathbb{R}_{>0}$ universal numerical constants and by $\mathfrak{C}(\cdot) \in \mathbb{R}_{>0}$ constants depending only on the argument. In both cases, the values of the constants may change from line to line. □

2.3. Nonparametric density estimation - known error distribution

In case of a known error distribution of U we follow [10] in defining a spectral cut-off density estimator $\tilde{f}_k$ of f for each $x \in \mathbb{R}_{>0}$ and tuning parameter $k \in \mathbb{R}_{>0}$ by

$$\tilde{f}_k(x) := \int_{[-k,k]} x^{-c-\iota 2\pi t} \widehat{\mathrm{M}}_Y(t)\, \mathrm{M}_U^\dagger(t) \mathrm{d}\lambda(t) = \mathrm{M}_c^{-1}\left[\widetilde{\mathrm{M}}_X^k\right](x). \tag{7}$$

assuming that $\mathbb{1}_{[-k,k]} \mathrm{M}_U^\dagger \in \mathbb{L}^2$ for each $k \in \mathbb{R}_{>0}$. The last condition ensures evidently that $\tilde{f}_k$ is well-defined. Note, in general the estimator $\tilde{f}_k$ does not define a probability density. For a simple solution to this problem we refer to [18], where a $\mathbb{L}^2$-projection onto the class of non-negative functions integrating to one is proposed in (3.1.15) on p. 63. If we require that $\mathrm{M}_U \neq 0$ a.s. then we have $f = \mathrm{M}_c^{-1}[\mathrm{M}_X] = \mathrm{M}_c^{-1}\left[\mathrm{M}_Y \mathrm{M}_U^\dagger\right]$ due to the Mellin inversion formula and the multiplicative convolution theorem. Intuitively speaking, from the last identity we obtain the estimator in (7) by truncating and replacing the unknown Mellin transformation M_Y of f^Y by its empirical counter part $\widehat{\mathrm{M}}_Y$. In [10] the $\mathbb{L}^2_+(\mathrm{x}^{2c-1})$-risk of $\tilde{f}_k$ is analysed, which for each $k \in \mathbb{R}_{>0}$ is written as

$$\mathbb{E}\left[\|\tilde{f}_k - f\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})}\right] = \mathbb{E}\left[\left\|\widetilde{\mathrm{M}}_X^k - \mathrm{M}_X\right\|^2_{\mathbb{L}^2}\right]$$
$$= \left\|\mathbb{1}_{[-k,k]^C} \mathrm{M}_X\right\|^2_{\mathbb{L}^2} + \frac{1}{n}\left\|\mathbb{1}_{[-k,k]} \mathbb{V}_Y \mathrm{M}_U^\dagger\right\|^2_{\mathbb{L}^2}, \quad (8)$$

where the first equality follows directly from Plancherel's identity (see (3)) and the second equality is due to the usual squared bias and variance decomposition. Let us emphasise that accessing the estimation accuracy by an integrated mean squared error, i.e. a $\mathbb{L}^2_+$-risk, corresponds in eq. (8) to a special case, namely choosing $c = \frac{1}{2}$. In this situation, for example, the identity in eq. (5) holds under the integrability and moment assumptions $f \in \mathbb{L}^2_+$ and $f^U \in \mathbb{L}^1_+(\mathrm{x}^{-\frac{1}{2}}) \cap \mathbb{L}^2_+$. However, allowing an arbitrary value $c \in \mathbb{R}$ makes the dependence on the developing point visible.

2.4. Nonparametric survival function estimation - known error distribution

In this paragraph, following [12] we recall an estimator of the survival function S^X of X based on the observation $\{Y_i\}_{i \in [\![n]\!]}$, when the error density f^U is known. The survival function of X satisfies

$$S^X : \begin{cases} \mathbb{R}_{>0} & \to [0,1] \\ x & \mapsto \mathbb{P}[X \geq x]. \end{cases}$$

As X admits the density f, we evidently for all $x \in \mathbb{R}$ have

$$S^X(x) = \int_{\mathbb{R}_{>0}} \mathbb{1}_{[x,\infty)}(y) f(y) \mathrm{d}\lambda_{>0}(y).$$

According to [12] we have $f \in \mathbb{L}^1_+(\mathrm{x}^{c-1}) \cap \mathbb{L}^2_+(\mathrm{x}^{2c-1})$ if and only if $S^X \in \mathbb{L}^1_+(\mathrm{x}^{c-2}) \cap \mathbb{L}^2_+(\mathrm{x}^{2c-3})$. Further, elementary computations show for each $t \in \mathbb{R}$ and $c \in \mathbb{R}_{>1}$ that

$$\mathrm{M}_{c-1}[S^X](t) = (c - 1 + \iota 2\pi t)^{-1} \mathrm{M}_X(t),$$

Exploiting the last identity [12] propose a spectral cut-off estimator of S^X given for all $k, x \in \mathbb{R}_{>0}$ by

$$\begin{aligned}\tilde{S}^X_k(x) &:= \int_{[-k,k]} x^{-c+1-\iota 2\pi t}\frac{\widehat{\mathrm{M}}_Y(t)\,\mathrm{M}^{\dagger}_U(t)}{c-1+\iota 2\pi t}\mathrm{d}\lambda(t)\\ &= \mathrm{M}^{-1}_{c-1}\left[(c-1+\iota 2\pi\cdot)^{-1}\widetilde{\mathrm{M}}^k_X(\cdot)\right](x).\end{aligned}$$

and study its $\mathbb{L}^2_+(\mathrm{x}^{2c-3})$-risk, which reads as

$$\begin{aligned}\mathbb{E}\left[\|\tilde{S}^X_k - S\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-3})}\right] &= \mathbb{E}\left[\left\|(c-1+\iota 2\pi\cdot)^{-1}(\widetilde{\mathrm{M}}^k_X - \mathrm{M}_X)\right\|^2_{\mathbb{L}^2}\right]\\ &= \mathbb{E}\left[\left\|\widetilde{\mathrm{M}}^k_X - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{t}_c)}\right]\\ &= \|\mathbb{1}_{[-k,k]^C}\,\mathrm{M}_X\|^2_{\mathbb{L}^2(\mathrm{t}_c)} + \frac{1}{n}\left\|\mathbb{1}_{[-k,k]}\mathbb{V}_Y\,\mathrm{M}^{\dagger}_U\right\|^2_{\mathbb{L}^2(\mathrm{t}_c)},\end{aligned} \tag{9}$$

where the Plancherel's identity (3) yields the first equality, the second makes use of the density function $\mathrm{t}_c(t) := \big((c-1)^2 + 4\pi^2 t^2\big)^{-1}$ for $t \in \mathbb{R}$ and the third states the squared bias and variance decomposition. As in Subsection 2.3 before, the analysis of the $\mathbb{L}^2_+$-estimation risk is covered as a special case, namely, by the choice $c = \frac{3}{2}$. In this situation, for example, the identity in eq. (5) holds under the integrability and moment assumptions $f \in \mathbb{L}^2_+(\mathrm{x}^2)$ and $f^U \in \mathbb{L}^1_+(\mathrm{x}^{\frac{1}{2}}) \cap \mathbb{L}^2_+(\mathrm{x}^2)$. Thus, again a general choice of c implies a more flexible risk analysis.

2.5. Oracle type inequality and minimax optimal rates

In the previous paragraphs, we have seen that the global $\mathbb{L}^2_+(\mathrm{x}^{2c-1})$-risk for $\tilde{f}_k$ in (8) and the global $\mathbb{L}^2_+(\mathrm{x}^{2c-3})$-risk for $\tilde{S}^X_k$ in (9) exactly equals

$$\mathbb{E}\left[\left\|\widetilde{\mathrm{M}}^k_X - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right] = \|\mathbb{1}_{[-k,k]^C}\,\mathrm{M}_X\|^2_{\mathbb{L}^2(\mathrm{v})} + \frac{1}{n}\left\|\mathbb{1}_{[-k,k]}\mathbb{V}_Y\,\mathrm{M}^{\dagger}_U\right\|^2_{\mathbb{L}^2(\mathrm{v})}, \tag{10}$$

with $\mathrm{v} = \mathbb{1}$ and $\mathrm{v} = \mathrm{t}_c$ accordingly. Therefore we study in the sequel the risk in (10) with an arbitrary density function v more in detail and consider oracle-type inequalities as well as minimax-optimal convergences rates. Let us summarise the assumptions we have so far imposed.

Assumption A.I.
Let X and U be independent and $\mathbb{R}_{>0}$-valued random variables with density f and f^U, respectively. Consider i.i.d. observations $\{Y_i\}_{i\in[\![n]\!]}$ following the law of $Y = X \cdot U$, which admits the density $f^Y = f \circledast f^U$. In addition, let

i) $f^U \in \mathbb{L}^1_+(\mathrm{x}^{c-1}) \cap \mathbb{L}^2_+(\mathrm{x}^{2c-1})$, $f \in \mathbb{L}^2_+(\mathrm{x}^{2c-1})$ and $f^Y \in \mathbb{L}^1_+(\mathrm{x}^{2(c-1)})$, set $\sigma^2_Y := 1 + \mathbb{E}[Y^{2(c-1)}] = 1 + \|f^Y\|_{\mathbb{L}^1_+(\mathrm{x}^{2(c-1)})}$.

ii) $\mathrm{v} : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be a (measurable) density function and set $\mathrm{v}_U := |\,\mathrm{M}^{\dagger}_U\,|^2\mathrm{v}$.

iii) $\mathrm{M}_X \in \mathbb{L}^2(\mathrm{v})$, $\mathrm{M}_U \neq 0$ a.s. and $\mathbb{1}_{[-k,k]} \in \mathbb{L}^2(\mathrm{v}_U)$ (hence $\mathbb{1}_{[-k,k]} \in \mathbb{L}^2(\mathrm{v})$) for each $k \in \mathbb{R}_{>0}$. □

Remark 2.1. Let us briefly discuss the condition $\mathrm{M}_U \neq 0$ a.s. in Assumption A.I iii). The estimation procedure relies on the multiplication theorem eq. (5), which states a decomposition of M_Y into the product of M_X and M_U. Knowing M_Y, we can recover M_X only if $\mathrm{M}_U \neq 0$ (except on a Lebesgue-null set). However, if $\mathrm{M}_U \neq 0$ is not satisfied, one could obtain an estimator of M_X restricted to the subset $\{\mathrm{M}_U \neq 0\} \subseteq \mathbb{R}$. Alternatively, the condition $\mathrm{M}_U \neq 0$ could be fulfilled for a different value of the development point $c \in \mathbb{R}$. Changing the development point is similar in spirit to the approach presented in [6], where the estimation procedure within the additive deconvolution problem on the real line is analysed on a *strip* as a subset of the complex plane $\mathbb{C}$. We should emphasise that both – the restriction to $\{\mathrm{M}_U \neq 0\}$ and the change of the development point c, such that $\mathrm{M}_U \neq 0$, necessitates the knowledge of M_U and hence of the error density f^U. It is not evident how these approaches can be transferred to the completely unknown error distribution case. □

Corollary 2.1. *Under Assumption A.I, there exists an optimal tuning parameter $k_{\mathfrak{o}} \in \mathbb{R}_{>0}$, such that*

$$\begin{aligned}\mathbb{E}\left[\left\|\widetilde{\mathrm{M}}_X^{k_{\mathfrak{o}}} - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right] &= \inf_{k\in\mathbb{R}_{>0}}\left\{\left\|\mathbb{1}_{[-k,k]^C}\,\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})} + \frac{1}{n}\left\|\mathbb{1}_{[-k,k]}\mathbb{V}_Y\,\mathrm{M}_U^{\dagger}\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right\}\\ &\le \sigma_Y^2\cdot\inf_{k\in\mathbb{R}_{>0}}\left\{\left\|\mathbb{1}_{[-k,k]^C}\,\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})} + \frac{1}{n}\left\|\mathbb{1}_{[-k,k]}\,\mathrm{M}_U^{\dagger}\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right\}. \quad (11)\end{aligned}$$

Proof of Corollary 2.1. The claim follows from the elementary risk decomposition stated in (10) into two terms depending on $k \in \mathbb{R}_{>0}$ and observing that the first term decreases while the second one increases for an increasing k. □

Note, that the optimal tuning parameter $k_{\mathfrak{o}} \in \mathbb{R}_{>0}$ provided in Corollary 2.1 is not feasible as it depends directly on the unknown Mellin transformation M_X. Therefore $k_{\mathfrak{o}}$ is called an oracle and hence the upper bound in (11) is also called an oracle-type inequality. In the paragraph below, we revisit a data-driven selection method for k to avoid this lack of information. For this choice we also obtain a oracle-type inequality similar to (11) up to a different constant. We next briefly recall minimax-optimal convergences rates for the global $\mathbb{L}^2(\mathrm{v})$-risk as for instance derived in [10] and [12]. To do so, we impose regularity assumptions on the Mellin transformation M_X and M_U. Firstly, let us recall the definition of the Mellin-Sobolev space with regularity parameter $s \in \mathbb{R}_{>0}$, given by

$$\mathbb{W}^s := \left\{h \in \mathbb{L}^2(\mathrm{x}^{2c-1}) :\, \left\|\mathrm{t}_2^{-\frac{s}{2}}\,\mathrm{M}_c[h]\right\|_{\mathbb{L}^2} = \|\mathrm{M}_c[h]\|_{\mathbb{L}^2(\mathrm{t}_2^{-s})} < \infty\right\}.$$

Further, for some radius $L \in \mathbb{R}_{>0}$, we consider the associated Mellin-Sobolev ellipsoid, given by

$$\mathbb{W}^s(L) := \left\{h \in \mathbb{W}^s :\, \|\mathrm{M}_c[h]\|_{\mathbb{L}^2(\mathrm{t}_2^{-s})} \le L\right\}.$$

In the following brief discussion we assume an unknown density $f \in \mathbb{W}^s(L)$ of X, where the regularity s is specified below. Regarding the Mellin transformation M_U, we assume in the sequel its *ordinary smoothness* ($\mathfrak{o.s.}$) or *super smoothness* ($\mathfrak{s.s.}$), i.e. for some decay parameter $\gamma \in \mathbb{R}_{>0}$ and all $t \in \mathbb{R}$,

$$\exists \mathfrak{c}_l, \mathfrak{c}_u \in \mathbb{R}_{>0}: \ \mathfrak{c}_l \cdot \mathrm{t}_2^{\frac{\gamma}{2}}(t) \leq |\,\mathrm{M}_U(t)| \leq \mathfrak{c}_u \cdot \mathrm{t}_2^{\frac{\gamma}{2}}(t) \tag{$\mathfrak{o.s.}$}$$

and

$$\exists \mathfrak{c}_l, \mathfrak{c}_u \in \mathbb{R}_{>0}: \ \exp(-\mathfrak{c}_l \cdot |t|^{2\gamma}) \leq |\,\mathrm{M}_U(t)| \leq \exp(-\mathfrak{c}_u \cdot |t|^{2\gamma}), \tag{$\mathfrak{s.s.}$}$$

respectively. In order to discuss the convergences rates for the upper bound of the $\mathbb{L}^2(\mathrm{v})$-risk in (11) under these regularity assumptions, we restrict ourselves to the choice $\mathrm{v} = \mathrm{t}_c^{-a}$ for $a \in \mathbb{R}$, observing that $a = 0$ corresponds to the global risk for estimating the density f and $a = -1$ corresponds to estimating the survival function S^X of X as discussed in Subsections 2.3 and 2.4. In the following Corollary 2.2, we derive upper risk bounds provided an optimal choice of $k_{\mathfrak{o}}$.

Corollary 2.2. *Under Assumption A.I, let $f \in \mathbb{W}^s(L)$ for some fixed $L \in \mathbb{R}_{>0}$.*

1. *If M_U satisfies eq. ($\mathfrak{o.s.}$), choose $k_{\mathfrak{o}} := n^{\frac{1}{2\gamma+2s+1}}$. Then,*

$$\mathbb{E}\left[\left\|\widetilde{\mathrm{M}}_X^{k_{\mathfrak{o}}} - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{t}_c^{-a})}\right] \leq \mathfrak{C}(a, s, c, L, \mathfrak{c}_l, \sigma_Y^2) \cdot n^{-\frac{2(s-a)}{2\gamma+2s+1}} \tag{12}$$

   for any $a \in (-1/2 - \gamma, s)$.

2. *If M_U satisfies eq. ($\mathfrak{s.s.}$), choose $k_{\mathfrak{o}} := (\log n)^{\frac{1}{2\gamma}}$. Then,*

$$\mathbb{E}\left[\left\|\widetilde{\mathrm{M}}_X^{k_{\mathfrak{o}}} - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{t}_c^{-a})}\right] \leq \mathfrak{C}(a, s, c, L, \mathfrak{c}_l, \sigma_Y^2) \cdot (\log n)^{-\frac{s-a}{\gamma}} \tag{13}$$

   for any $a < s$.

Proof of Corollary 2.2. Having Corollary 2.1 in mind, we start with the bias expression, which does not depend on M_U. To be more precise, if M_X belongs to the Mellin-Sobolev ellipsoid $\mathbb{W}^s(L)$ with $a < s$ then for all $k \in \mathbb{R}_{>0}$

$$\left\|\mathbb{1}_{[-k,k]^C}\,\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{t}_c^{-a})} \leq L^2 \left\|\mathbb{1}_{[-k,k]^C}\mathrm{t}_2^s \mathrm{t}_c^{-a}\right\|_{\mathbb{L}^\infty} \leq \mathfrak{C}(c, a, s, L) k^{-2(s-a)}. \tag{14}$$

Regarding the variance expression, assume first that M_U is ordinary smooth, i.e. eq. ($\mathfrak{o.s.}$) holds, and $a + \gamma + \frac{1}{2} > 0$. Then for all $k \in \mathbb{R}_{>0}$

$$\left\|\mathbb{1}_{[-k,k]}\,\mathrm{M}_U^{\dagger}\right\|^2_{\mathbb{L}^2(\mathrm{t}_c^{-a})} \leq \mathfrak{c}_l \cdot \left\|\mathbb{1}_{[-k,k]}\mathrm{t}_2^{-\gamma}\mathrm{t}_c^{-a}\right\|_{\mathbb{L}^1} \leq \mathfrak{C}(c, a, s, \mathfrak{c}_l) k^{2(a+\gamma)+1}.$$

Selecting $k_{\mathfrak{o}} = n^{\frac{1}{2\gamma+2s+1}}$ and combining these two estimates together with eq. (11), we obtain eq. (12). Similarly, if M_U is super smooth, i.e. eq. ($\mathfrak{s.s.}$) holds, we have for any $a < s$ and $k \in \mathbb{R}_{>0}$

$$\left\|\mathbb{1}_{[-k,k]}\,\mathrm{M}_U^{\dagger}\right\|^2_{\mathbb{L}^2(\mathrm{v})} \leq \left\|\mathbb{1}_{[-k,k]}\exp(2\mathfrak{c}_l|\cdot|^{2\gamma})\mathrm{t}_c^{-a}\right\|_{\mathbb{L}^1} \leq \mathfrak{C}(c,a,s,\mathfrak{c}_l)k^{(1-2(\gamma-a))_+}\exp(k^{2\gamma}).$$

Selecting $k_{\mathfrak{o}} = (\log n)^{\frac{1}{2\gamma}}$ and combining the last bound and eq. (14), we obtain eq. (13), which completes the proof. □

Remark 2.2. Let us briefly discuss the choices of $a=0$ and $a=-1$, which provide upper bounds for the global estimation risk for the density f and the survival function S^X, respectively:

1. Density estimation: Choosing $a=0$ for all $f\in\mathbb{W}^s(L)$ leads to
$$\mathbb{E}\left[\|\tilde{f}_{k_{\mathfrak{o}}}-f\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})}\right] \lesssim n^{-\frac{2s}{2\gamma+2s+1}}$$
if M_U is ordinary smooth, and
$$\mathbb{E}\left[\|\tilde{f}_{k_{\mathfrak{o}}}-f\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})}\right] \lesssim (\log n)^{-\frac{s}{\gamma}}$$
if M_U is super smooth. In [10] it is shown that these rates are minimax-optimal for $n\to\infty$.
2. Survival function estimation: Choosing $a=-1$ for all $f\in\mathbb{W}^s(L)$ leads to
$$\mathbb{E}\left[\|\tilde{S}^X_{k_{\mathfrak{o}}}-S\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-3})}\right] \lesssim n^{-\frac{2(s+1)}{2\gamma+2s+1}}$$
if M_U is ordinary smooth, and
$$\mathbb{E}\left[\|\tilde{S}^X_{k_{\mathfrak{o}}}-S\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-3})}\right] \lesssim (\log n)^{-\frac{s+1}{\gamma}}$$
if M_U is super smooth. In [12] it is shown that these rates are minimax-optimal for $n\to\infty$. □

2.6. Data driven estimation

[10] and [12] propose a data-driven choice of the tuning parameter k by model selection exploiting the theory of [3], where we refer to [23] for an extensive overview. More precisely they consider

$$\tilde{k} :\in \operatorname*{arg\,min}_{k\in\mathcal{K}}\left\{-\left\|\mathbb{1}_{[-k,k]}\widehat{\mathrm{M}}_Y\,\mathrm{M}_U^{\dagger}\right\|^2_{\mathbb{L}^2(\mathrm{v})}+\mathrm{pen}_k\right\} = \operatorname*{arg\,min}_{k\in\mathcal{K}}\left\{-\left\|\widetilde{\mathrm{M}}_X^k\right\|^2_{\mathbb{L}^2(\mathrm{v})}+\mathrm{pen}_k\right\}, \quad (15)$$

where $\{\mathrm{pen}_k\}_{k\in\mathbb{R}_{>0}}$ is a family of penalties and $\mathcal{K}\subset\mathbb{R}_{>0}$ is an appropriate finite subset specified later. The aim is to analyse the $\mathbb{L}^2(\mathrm{v})$-risk, namely

$$\mathbb{E}\left[\left\|\widetilde{\mathrm{M}}_X^{\tilde{k}}-\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right].$$

In the sequel we introduce a family of penalties and set of models which differ from the original works in preparation of the procedure presented in Sections 3 and 4 below. For the upper risk bound we impose slightly stronger assumptions than Assumption A.I that we state next.

Assumption A.II.
In addition to Assumption A.I let $f^Y \in \mathbb{L}^1_+(\mathrm{x}^{8(c-1)}) \cap \mathbb{L}^\infty_+(\mathrm{x}^{2c-1})$ such that there exists $\eta_Y \in \mathbb{R}_{\geq 1}$ satisfying $\eta_Y \geq \max\left\{\|f^Y\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})}, \|f^Y\|_{\mathbb{L}^1_+(\mathrm{x}^{8(c-1)})}\right\}$. We set $a_Y := 6\|f^Y\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})}/\sigma_Y^2$ and $k_Y := 1 \vee 3a_Y^2$. □

Text-book computations as they can be found for instance in [14] are leading to the following standard key argument, which for any $k_\circ \in \mathcal{K}$ states

$$\left\|\widetilde{\mathrm{M}}_X^{\tilde{k}} - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})} \leq 3\left\|\mathbb{1}_{[-k_\circ,k_\circ]^C}\,\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})} + 4\,\mathrm{pen}_{k_\circ} + 8\max_{k\in\mathcal{K}}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\,\mathrm{M}_U^\dagger\right\|^2_{\mathbb{L}^2(\mathrm{v})} - \frac{\mathrm{pen}_k}{4}\right)_+\right\}, \quad (16)$$

using the shorthand notation $a_+ := \max\{0, a\}$ for any $a \in \mathbb{R}$. Recalling $\mathrm{v}_U := |\,\mathrm{M}_U^\dagger\,|^2\mathrm{v}$, the last summand in (16) reads as

$$\max_{k\in\mathcal{K}}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\,\mathrm{M}_U^\dagger\right\|^2_{\mathbb{L}^2(\mathrm{v})} - \frac{\mathrm{pen}_k}{4}\right)_+\right\} = \max_{k\in\mathcal{K}}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\right\|^2_{\mathbb{L}^2(\mathrm{v}_U)} - \frac{\mathrm{pen}_k}{4}\right)_+\right\}.$$

Taking the expectation in the last display, the next proposition allows to control its value. In its proof we make use of Lemma A.2 in Appendix A, which is based on a Talagrand inequality, stated in Lemma A.1. Here and subsequently, for an arbitrary density function $\mathrm{w} : \mathbb{R} \to \mathbb{R}_{\geq 0}$ satisfying $\mathbb{1}_{[-k,k]} \in \mathbb{L}^\infty(\mathrm{w})$ for each $k \in \mathbb{R}_{>0}$, we denote by

$$\Delta_k^{\mathrm{w}} := \left\|\mathbb{1}_{[-k,k]}\right\|_{\mathbb{L}^\infty(\mathrm{w})} \text{ and } \delta_k^{\mathrm{w}} := \frac{\log(\Delta_k^{\mathrm{w}} \vee (k+2))}{\log(k+2)} \in \mathbb{R}_{\geq 1}, \ \forall k \in \mathbb{R}_{\geq 1}. \quad (17)$$

Proposition 2.3 (Concentration inequality)**.** *Under Assumption A.II define* $\Delta_k^{\mathrm{v}_U}$ *and* $\delta_k^{\mathrm{v}_U}$ *as in* (17) *with* $\mathrm{v}_U := |\,\mathrm{M}_U^\dagger\,|^2\mathrm{v}$. *For* $n \in \mathbb{N}$ *consider*

$$k_n := \max\{k \in [\![n^2]\!] : k\Delta_k^{\mathrm{v}_U} \leq n^2\Delta_1^{\mathrm{v}_U}\}.$$

We then have

$$\mathbb{E}\left[\max_{k\in[\![k_n]\!]}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\right\|^2_{\mathbb{L}^2(\mathrm{v}_U)} - 12\sigma_Y^2\Delta_k^{\mathrm{v}_U}\delta_k^{\mathrm{v}_U}kn^{-1}\right)_+\right\}\right] \leq \mathfrak{C}\cdot\eta_Y(1 \vee k_Y\Delta_{k_Y}^{\mathrm{v}_U})\cdot n^{-1}.$$

Having Proposition 2.3 at hand we set $\mathcal{K} := [\![k_n]\!] \subset \mathbb{N}$ and define for each $k \in [\![k_n]\!]$

$$\mathrm{pen}_k^{\mathrm{v}_U} := 48\Delta_k^{\mathrm{v}_U}\delta_k^{\mathrm{v}_U} kn^{-1}.$$

Evidently, choosing $\mathrm{pen}_k := \sigma_Y^2\,\mathrm{pen}_k^{\mathrm{v}_U}$ Proposition 2.3 allows to bound the expectation of the last summand in (16). Unfortunately, $\sigma_Y^2 = 1 + \mathbb{E}[Y^{2(c-1)}]$ is unknown to us. However, we have at our disposal an unbiased estimator given by

$$\widehat{\sigma}_Y^2 := 1 + n^{-1} \sum_{i\in[\![n]\!]} Y_i^{2(c-1)}.$$

Hence, replacing subsequently the unknown σ_Y^2 by its empirical counterpart $\widehat{\sigma}_Y^2$ we consider the data driven choice

$$\hat{k} :\in \operatorname*{arg\,min}_{k\in[\![k_n]\!]} \left\{ -\left\|\widetilde{\mathrm{M}}_X^k\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 2\widehat{\sigma}_Y^2\,\mathrm{pen}_k^{\mathrm{v}_U} \right\}. \tag{18}$$

Corollary 2.4. *Under Assumption A.II, we have*

$$\mathbb{E}\left[\left\|\widetilde{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2\right] \le 6\sigma_Y^2 \cdot \min_{k\in[\![k_n]\!]} \left\{ \left\|\mathbb{1}_{[-k,k]^C}\,\mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + \Delta_k^{\mathrm{v}_U}\delta_k^{\mathrm{v}_U} kn^{-1} \right\} + \mathfrak{C}\cdot \eta_Y(1\vee k_Y\Delta_{k_Y}^{\mathrm{v}_U})\cdot n^{-1}. \tag{19}$$

Proof of Corollary 2.4. Similar computations as presented in [10] show a slightly changed version of the key argument (16), which reads for each $k_\circ \in [\![k_n]\!]$ as

$$\begin{aligned}\left\|\widetilde{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2 &\le 3\left\|\mathbb{1}_{[-k_\circ,k_\circ]^C}\,\mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 2\sigma_Y^2\,\mathrm{pen}_{k_\circ}^{\mathrm{v}_U} \\ &\quad + 4\widehat{\sigma}_Y^2\,\mathrm{pen}_{k_\circ}^{\mathrm{v}_U} + 2(\sigma_Y^2 - 2\widehat{\sigma}_Y^2)_+\,\mathrm{pen}_{k_n}^{\mathrm{v}_U} \\ &\quad + 8\max_{k\in[\![k_n]\!]}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\right\|_{\mathbb{L}^2(\mathrm{v}_U)}^2 - \frac{\sigma_Y^2\,\mathrm{pen}_k^{\mathrm{v}_U}}{4}\right)_+\right\}.\end{aligned} \tag{20}$$

By applying the expectation on both sides and taking

$$k_\circ :\in \operatorname*{arg\,min}_{k\in[\![k_n]\!]} \left\{ \left\|\mathbb{1}_{[-k,k]^C}\,\mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + \Delta_k^{\mathrm{v}_U}\delta_k^{\mathrm{v}_U} kn^{-1} \right\}, \tag{21}$$

we immediately obtain the claim eq. (19) due to $\mathbb{E}[\widehat{\sigma}_Y^2] = \sigma_Y^2$, Proposition 2.3 as well as

$$\mathbb{E}\left[\left(\frac{\sigma_Y^2}{2} - \widehat{\sigma}_Y^2\right)_+ \mathrm{pen}_{k_n}^{\mathrm{v}_U}\right] \le \mathfrak{C}\eta_Y\Delta_1^{\mathrm{v}_U} n^{-1}.$$

This completes the proof. □

At this point we want to stress out again that specifying $\mathrm{v} = \mathbb{1}$ and $\mathrm{v} = \mathrm{t}_c$, respectively, we obtain directly from Corollary 2.4 upper bounds for

$$\mathbb{E}\left[\|\tilde{f}_{\hat{k}} - f\|_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})}^2\right] \text{ and } \mathbb{E}\left[\|\tilde{S}_{\hat{k}}^X - S^X\|_{\mathbb{L}^2_+(\mathrm{x}^{3c-2})}^2\right]$$

due to Plancherel's identity (3). Similar to Subsection 2.5 we assume in the following brief discussion a density $f \in \mathbb{W}^s(L)$ of X, where the regularity $s \in \mathbb{R}_{>0}$ is specified below. Regarding the Mellin transformation M_U, we subsequently assume again its *ordinary smoothness* ($\mathfrak{o.s.}$) or *super smoothness* ($\mathfrak{s.s.}$). In order to discuss the convergences rates for the upper bound of the $\mathbb{L}^2(\mathrm{v})$-risk in Corollary 2.4 under these regularity assumptions, we restrict ourselves to the choice $\mathrm{v} := \mathrm{t}_c^{-a}$ for $a \in \mathbb{R}$, observing that $a = 0$ and $a = -1$ correspond to the global risk for estimating the density f and the survival function S^X of X, respectively.

Corollary 2.5. *Under Assumption A.II, let $f \in \mathbb{W}^s(L)$ for some fixed $L \in \mathbb{R}_{>0}$.*

1. *If M_U satisfies eq. ($\mathfrak{o.s.}$), then,*

$$\mathbb{E}\left[\left\|\widetilde{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{t}_c^{-a})}\right] \leq \mathfrak{C}(a, \gamma, s, c, L, \mathfrak{c}_l, \eta_Y, k_Y, \sigma_Y^2) \cdot n^{-\frac{2(s-a)}{2\gamma+2s+1}}$$

   for any $a \in (-1/2 - \gamma, s)$.

2. *If M_U satisfies eq. ($\mathfrak{s.s.}$), then,*

$$\mathbb{E}\left[\left\|\widetilde{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{t}_c^{-a})}\right] \leq \mathfrak{C}(a, \gamma, s, c, L, \mathfrak{c}_l, \eta_Y, k_Y, \sigma_Y^2) \cdot (\log n)^{-\frac{s-a}{\gamma}}$$

   for any $a < s$.

Proof of Corollary 2.5. Starting with Corollary 2.4, we use again the bound eq. (14) as the bias expression is the same as in Corollary 2.2. However the variance expression differ compared to Corollary 2.2. Assume first that M_U is ordinary smooth, i.e. eq. ($\mathfrak{o.s.}$) holds, and $a + \gamma + \frac{1}{2} > 0$. Then for all $k \in \mathbb{R}_{>0}$

$$\Delta_k^{\mathrm{v}_U} = \left\|\mathbb{1}_{[-k,k]}|\,\mathrm{M}_U^{\dagger}\,|^2\right\|_{\mathbb{L}^\infty(\mathrm{v})} \leq \mathfrak{c}_l \cdot \left\|\mathbb{1}_{[-k,k]}\mathrm{t}_2^{-\gamma}\mathrm{t}_c^{-a}\right\|_{\mathbb{L}^\infty} \leq \mathfrak{C}(c, a, s, \mathfrak{c}_l) k^{2(a+\gamma)}.$$

Consequently, it follows $\delta_k^{\mathrm{v}_U} \leq \mathfrak{C}(c, a, s, \mathfrak{c}_l)$ and hence,

$$\Delta_k^{\mathrm{v}_U}\delta_k^{\mathrm{v}_U} k \leq \mathfrak{C}(c, a, s, \mathfrak{c}_l) k^{2(a+\gamma)+1}. \tag{22}$$

Selecting $k_{\mathfrak{o}} = \lfloor n^{\frac{1}{2\gamma+2s+1}} \rfloor \in [\![k_n]\!]$ and combining these two estimates together with eq. (19), we obtain

$$\min_{k \in [\![k_n]\!]}\left\{\left\|\mathbb{1}_{[-k,k]^C}\,\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})} + \Delta_k^{\mathrm{v}_U}\delta_k^{\mathrm{v}_U} k n^{-1}\right\} \leq \mathfrak{C}(c, a, s, \mathfrak{c}_l) \cdot n^{-\frac{2(s-a)}{2\gamma+2s+1}} \tag{23}$$

and thus with Corollary 2.4 follows the claim. Similarly, if M_U is super smooth, i.e. eq. ($\mathfrak{s.s.}$) holds, we have for any $a < s$ and $k \in \mathbb{R}_{>0}$

$$\begin{aligned}\Delta_k^{\mathrm{v}_U} = \left\|\mathbb{1}_{[-k,k]}|\,\mathrm{M}_U^{\dagger}\,|^2\right\|_{\mathbb{L}^\infty(\mathrm{v})} &\leq \left\|\mathbb{1}_{[-k,k]}\exp(2\mathfrak{c}_l|\cdot|^{2\gamma})\mathrm{t}_c^{-a}\right\|_{\mathbb{L}^\infty} \\ &\leq \mathfrak{C}(c, a, s, \mathfrak{c}_l) k^{2(a)_+}\exp(k^{2\gamma}).\end{aligned}$$

Consequently, it follows $\delta_k^{\mathrm{v}_U} \le \mathfrak{C}(c,a,s,\mathfrak{c}_l)k^{2\gamma}$ and hence,

$$\Delta_k^{\mathrm{v}_U}\delta_k^{\mathrm{v}_U}k \le \mathfrak{C}(c,a,s,\mathfrak{c}_l)k^{2(a)_+ + 2\gamma+1}\exp(k^{2\gamma}). \tag{24}$$

Selecting $k_{\mathfrak{o}} = \lfloor (\log n)^{\frac{1}{2\gamma}} \rfloor \in [\![k_n]\!]$ and combining the last bound and eq. (19), we obtain

$$\min_{k\in[\![k_n]\!]}\left\{\left\|\mathbb{1}_{[-k,k]^C}\,\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})} + \Delta_k^{\mathrm{v}_U}\delta_k^{\mathrm{v}_U}kn^{-1}\right\} \le \mathfrak{C}(c,a,s,\mathfrak{c}_l)\cdot(\log n)^{-\frac{s-a}{\gamma}} \tag{25}$$

and thus, with Corollary 2.4 the second claim, which completes the proof. □

Remark 2.3. Let us briefly revisit the choices of $a = 0$ and $a = -1$, which provide upper bounds for the global estimation risk for the density f and the survival function S^X, respectively:

1. Density estimation: Choosing $a = 0$ for all $f \in \mathbb{W}^s(L)$ leads to

$$\mathbb{E}\left[\left\|\tilde{f}_{\hat{k}} - f\right\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})}\right] \lesssim n^{-\frac{2s}{2\gamma+2s+1}}$$

if M_U is ordinary smooth, and

$$\mathbb{E}\left[\left\|\tilde{f}_{\hat{k}} - f\right\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})}\right] \lesssim (\log n)^{-\frac{s}{\gamma}}$$

if M_U is super smooth. The rates coincide with the minimax-optimal rates presented in [10] (compare Corollary 2.2).

2. Survival function estimation: Choosing $a = -1$ for all $f \in \mathbb{W}^s(L)$ leads to

$$\mathbb{E}\left[\left\|\tilde{S}^X_{\hat{k}} - S\right\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-3})}\right] \lesssim n^{-\frac{2(s+1)}{2\gamma+2s+1}}$$

if M_U is ordinary smooth, and

$$\mathbb{E}\left[\left\|\tilde{S}^X_{\hat{k}} - S\right\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-3})}\right] \lesssim (\log n)^{-\frac{s+1}{\gamma}}$$

if M_U is super smooth. The rates coincide with the minimax-optimal rates presented in [12] (compare Corollary 2.2).

We should emphasise that comparing the rates in Corollaries 2.2 and 2.5 there is no additional price for adaptation to pay. □

3. Estimation strategy for unknown error density

After recapitulating an estimation strategy for the multiplicative deconvolution problem assuming the error density is known in advance, we dismiss this assumption in this section. Inspired by similar ideas for additive deconvolution problems (see for instance [25] and [20]), we study estimation in the multiplicative deconvolution problem with unknown error density. In addition to i.i.d. observations $\{Y_i\}_{i\in[\![n]\!]}$ following the law of the multiplicative measurement model $Y = X\cdot U$, we have access to additional measurements $\{U_j\}_{j\in[\![m]\!]}$, $m \in \mathbb{N}$, which are i.i.d. drawn following the law of U independently of the first sample $\{Y_i\}_{i\in[\![n]\!]}$. We estimate the Mellin transformation M_U by its empirical counterpart given for each $t \in \mathbb{R}$ by

$$\widehat{\mathrm{M}}_U(t) := m^{-1} \sum_{j\in[\![m]\!]} U_j^{c-1+\iota 2\pi t}.$$

Similarly to $\widehat{\mathrm{M}}_Y$, we observe that $\mathbb{E}[\widehat{\mathrm{M}}_U(t)] = \mathrm{M}_U(t)$ for all $t \in \mathbb{R}$, i.e. $\widehat{\mathrm{M}}_U$ is an unbiased estimator of M_U. Here and subsequently, the expectation $\mathbb{E}$ is considered with respect to the joint distribution of $\{Y_i\}_{i\in[\![n]\!]}$ and $\{U_j\}_{j\in[\![m]\!]}$. As in Section 2 we intend to divide by $\widehat{\mathrm{M}}_U$ whenever it is well defined, or in equal multiply with $\widehat{\mathrm{M}}_U^\dagger = \frac{1}{\widehat{\mathrm{M}}_U}\mathbb{1}_{\{\widehat{\mathrm{M}}_U\neq 0\}}$ (see Notation 2.1). Note that the indicator set $\{\widehat{\mathrm{M}}_U \neq 0\} := \{t \in \mathbb{R} : \widehat{\mathrm{M}}_U(t) \neq 0\}$ is not deterministic anymore since it depends on the random variables $\{U_j\}_{j\in[\![m]\!]}$. However, we note that $|\widehat{\mathrm{M}}_U^\dagger|$ is generally unbounded on the event $\{\widehat{\mathrm{M}}_U \neq 0\}$ which would lead to an unstable estimation. Hence, we truncate $\widehat{\mathrm{M}}_U$ sufficiently far away from zero. Recalling that n and m denote the samples sizes of $\{Y_i\}_{i\in[\![n]\!]}$ and $\{U_j\}_{j\in[\![m]\!]}$, we thus define in accordance with [25] the random indicator set

$$\mathfrak{M} := \left\{(m\wedge n)|\widehat{\mathrm{M}}_U|^2 \geq 1\right\} := \left\{t \in \mathbb{R} : (m\wedge n)|\widehat{\mathrm{M}}_U(t)|^2 \geq 1\right\},$$

which only depends on the additional measurements $\{U_j\}_{j\in[\![m]\!]}$. Similarly as in Section 2 we have now all ingredients to define an estimator of the unknown Mellin transformation M_X. Indeed, having the convolution theorem (5) in mind (stating $\mathrm{M}_Y = \mathrm{M}_X \cdot \mathrm{M}_U$) we propose as an estimator of M_X

$$\widehat{\mathrm{M}}_X := \widehat{\mathrm{M}}_Y \widehat{\mathrm{M}}_U^\dagger \mathbb{1}_{\mathfrak{M}}.$$

To simplify the presentation later, we further write $\widehat{\mathrm{M}}_X^k := \widehat{\mathrm{M}}_X \mathbb{1}_{[-k,k]}$ and $\mathrm{M}_X^k := \mathrm{M}_X \mathbb{1}_{[-k,k]}$ for any $k \in \mathbb{R}_{>0}$.

3.1. Nonparametric density estimation - unknown error density

Motivated by the estimation strategy in Subsection 2.3 we propose in this paragraph a thresholded spectral cut-off density estimator for f in the multiplicative deconvolution problem with unknown error distribution.

Definition 3.1 (Thresholded spectral cut-off estimator)**.** Assuming an unknown density $f \in \mathbb{L}^2_+(\mathrm{x}^{2c-1})$ of X for some $c \in \mathbb{R}$ define the thresholded spectral density estimator $\hat{f}_k$ for each $k, x \in \mathbb{R}_{>0}$ by

$$\hat{f}_k(x) := \int_{[-k,k]} x^{-c-\iota 2\pi t}\widehat{\mathrm{M}}_Y(t)\widehat{\mathrm{M}}_U^\dagger(t)\mathbb{1}_{\mathfrak{M}}(t)\mathrm{d}\lambda(t) = \mathrm{M}_c^{-1}\left[\widehat{\mathrm{M}}_X^k\right](x).$$

The next lemma provides a representation of the global $\mathbb{L}^2_+(\mathrm{x}^{2c-1})$-risk of the estimator $\hat{f}_k$ similar to the decomposition (8) of the global $\mathbb{L}^2_+(\mathrm{x}^{2c-1})$-risk of $\tilde{f}_k$ in Subsection 2.3.

Lemma 3.2 (Risk representation)**.** *For $k \in \mathbb{R}_{>0}$ consider the density estimator $\hat{f}_k$ given in Definition 3.1 and recall the definition of $\mathbb{V}^2_Y$ (see Notation 2.1). We then have*

$$\begin{aligned}\mathbb{E}\left[\left\|\hat{f}_k - f\right\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})}\right] &= \mathbb{E}\left[\left\|\widehat{\mathrm{M}}^k_X - \mathrm{M}_X\right\|^2_{\mathbb{L}^2}\right] \\ &= \left\|\mathrm{M}_X \mathbb{1}_{[-k,k]^C}\right\|^2_{\mathbb{L}^2} + \frac{1}{n}\mathbb{E}\left[\left\|\widehat{\mathrm{M}}^\dagger_U \mathbb{1}_{\mathfrak{M}} \mathbb{V}_Y \mathbb{1}_{[-k,k]}\right\|^2_{\mathbb{L}^2}\right] \\ &\quad + \mathbb{E}\left[\left\|\mathrm{M}^k_X \mathbb{1}_{\mathfrak{M}^C}\right\|^2_{\mathbb{L}^2}\right] + \mathbb{E}\left[\left\|\widehat{\mathrm{M}}^\dagger_U \mathbb{1}_{\mathfrak{M}} (\mathrm{M}_U - \widehat{\mathrm{M}}_U)\,\mathrm{M}^k_X\right\|^2_{\mathbb{L}^2}\right].\end{aligned}$$

Proof of Lemma 3.2. Recalling the definition of $\hat{f}_k$ as well as Plancherel's identity (see eq. (3)), we have

$$\begin{aligned}\left\|\hat{f}_k - f\right\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})} &= \left\|\widehat{\mathrm{M}}^k_X - \mathrm{M}_X\right\|^2_{\mathbb{L}^2} = \left\|\widehat{\mathrm{M}}^k_X - \mathrm{M}_X(\mathbb{1}_{[-k,k]} + \mathbb{1}_{[-k,k]^C})\right\|^2_{\mathbb{L}^2} \\ &= \left\|\mathrm{M}_X \mathbb{1}_{[-k,k]} - \mathbb{1}_{[-k,k]}\widehat{\mathrm{M}}_Y \widehat{\mathrm{M}}^\dagger_U \mathbb{1}_{\mathfrak{M}}\right\|^2_{\mathbb{L}^2} + \left\|\mathrm{M}_X \mathbb{1}_{[-k,k]^C}\right\|^2_{\mathbb{L}^2} \\ &= \left\|\widehat{\mathrm{M}}^\dagger_U \mathbb{1}_{\mathfrak{M}}(\widehat{\mathrm{M}}_Y - \widehat{\mathrm{M}}_U \mathrm{M}_X)\mathbb{1}_{[-k,k]}\right\|^2_{\mathbb{L}^2} \\ &\quad + \left\|\mathrm{M}_X \mathbb{1}_{[-k,k]}\mathbb{1}_{\mathfrak{M}^C}\right\|^2_{\mathbb{L}^2} + \left\|\mathrm{M}_X \mathbb{1}_{[-k,k]^C}\right\|^2_{\mathbb{L}^2},\end{aligned}$$

where we used in the last step that

$$\begin{aligned}\mathrm{M}_X \mathbb{1}_{[-k,k]} &= \mathrm{M}_X \mathbb{1}_{[-k,k]}(\mathbb{1}_{\mathfrak{M}} + \mathbb{1}_{\mathfrak{M}^C}) \\ &= \widehat{\mathrm{M}}^\dagger_U \widehat{\mathrm{M}}_U \mathbb{1}_{\mathfrak{M}}\,\mathrm{M}_X \mathbb{1}_{[-k,k]} + \mathrm{M}_X \mathbb{1}_{[-k,k]}\mathbb{1}_{\mathfrak{M}^C}.\end{aligned}$$

Studying only the first summand further we obtain for each $t \in \mathbb{R}$

$$n\mathbb{E}\left[|\widehat{\mathrm{M}}_Y(t) - \mathrm{M}_Y(t)|^2\right] = \mathbb{E}\left[|Y_1^{c-1+\iota 2\pi t} - \mathrm{M}_Y(t)|^2\right] = \mathbb{V}^2_Y(t).$$

By exploiting the independence of $\{Y_i\}_{i\in[\![n]\!]}$ and $\{U_j\}_{j\in[\![m]\!]}$, we finally have

$$\begin{aligned}\mathbb{E}\left[\left\|\widehat{\mathrm{M}}^\dagger_U \mathbb{1}_{\mathfrak{M}}(\widehat{\mathrm{M}}_Y - \widehat{\mathrm{M}}_U \mathrm{M}_X)\mathbb{1}_{[-k,k]}\right\|^2_{\mathbb{L}^2}\right] &= \mathbb{E}\left[\left\|\widehat{\mathrm{M}}^\dagger_U \mathbb{1}_{\mathfrak{M}}(\mathrm{M}_U - \widehat{\mathrm{M}}_U)\,\mathrm{M}^k_X\right\|^2_{\mathbb{L}^2}\right] \\ &\quad + \frac{1}{n}\mathbb{E}\left[\left\|\widehat{\mathrm{M}}^\dagger_U \mathbb{1}_{\mathfrak{M}} \mathbb{V}_Y \mathbb{1}_{[-k,k]}\right\|^2_{\mathbb{L}^2}\right],\end{aligned}$$

which shows the claim. □

At this point we want to stress out that the $\mathbb{L}^2_+(\mathrm{x}^{2c-1})$-risk representation of Lemma 3.2 for $\hat{f}_k$ has a very similar structure as the corresponding risk representation of $\tilde{f}_k$ in (8) assuming the error density f^U to be known. Indeed, the first term remains the same – it represents the bias term, which can be specified later, considering different regularity assumptions on f. Later, we see that the second term actually represents the variance term. The two additional summands, only depending on the additional measurements $\{U_j\}_{j\in[\![m]\!]}$, occur in this particular situation, where we estimate the unknown M_U as well.

3.2. Nonparametric survival analysis - unknown error density

Considering the survival analysis in Subsection 2.4, we propose now an estimator for the survival function S^X of X under multiplicative measurement errors with unknown error density f^U and additional measurements $\{U_j\}_{j\in[\![m]\!]}$. Indeed, we follow the definition of $\tilde{S}_k$ and analogously as for the density estimator $\tilde{f}_k$ we replace the unknown Mellin transformation M_U by its (sufficiently truncated) counterpart, which leads to the following definition.

Definition 3.3 (Thresholded spectral cut-off estimator of S^X)**.** Assuming an unknown density $f \in \mathbb{L}^2_+(\mathrm{x}^{2c-1})$ of X for some $c \in \mathbb{R}_{>1}$. The thresholded spectral cut-off estimator $\widehat{S}^X_k$ of the survival function S^X of X is defined for each $k, x \in \mathbb{R}_{>0}$ by

$$\begin{aligned}\widehat{S}^X_k(x) &:= \int_{[-k,k]} x^{-c+1-\iota 2\pi t}\frac{\widehat{\mathrm{M}}_Y(t)\widehat{\mathrm{M}}^{\dagger}_U(t)}{c-1+\iota 2\pi t}\mathbb{1}_{\mathfrak{M}}(t)\mathrm{d}\lambda(t)\\ &= \mathrm{M}^{-1}_{c-1}\left[(c-1+\iota 2\pi\cdot)^{-1}\widehat{\mathrm{M}}^k_X\right](x).\end{aligned}$$

Similar to (9) we are again interested in quantifying the accuracy of $\widehat{S}^X_k$ in terms of its global $\mathbb{L}^2_+(\mathrm{x}^{2c-3})$-risk. The representation in the next corollary follows line by line the proof of Lemma 3.2 and is hence omitted.

Corollary 3.4 (Risk Representation)**.** *For $k \in \mathbb{R}_{>0}$ consider the estimator $\hat{S}_k$ given in Definition 3.3 and recall that $\mathrm{t}_c : \mathbb{R} \to \mathbb{R}_{>0}$ with $t \mapsto \mathrm{t}_c(t) := \big((c-1)^2 + 4\pi^2 t^2\big)^{-1}$. We then have*

$$\begin{aligned}\mathbb{E}\left[\left\|\widehat{S}^X_k - S^X\right\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-3})}\right] &= \mathbb{E}\left[\left\|(c-1+\iota 2\pi\cdot)^{-1}\left(\widehat{\mathrm{M}}^k_X - \mathrm{M}_X\right)\right\|^2_{\mathbb{L}^2}\right]\\ &= \mathbb{E}\left[\left\|\widehat{\mathrm{M}}^k_X - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{t}_c)}\right]\\ &= \left\|\mathrm{M}_X\mathbb{1}_{[-k,k]^C}\right\|^2_{\mathbb{L}^2(\mathrm{t}_c)} + \frac{1}{n}\mathbb{E}\left[\left\|\widehat{\mathrm{M}}^{\dagger}_U\mathbb{1}_{\mathfrak{M}}\mathbb{V}_Y\mathbb{1}_{[-k,k]}\right\|^2_{\mathbb{L}^2(\mathrm{t}_c)}\right]\\ &\quad + \mathbb{E}\left[\left\|\mathrm{M}^k_X\mathbb{1}_{\mathfrak{M}^C}\right\|^2_{\mathbb{L}^2(\mathrm{t}_c)}\right] + \mathbb{E}\left[\left\|\widehat{\mathrm{M}}^{\dagger}_U\mathbb{1}_{\mathfrak{M}}(\mathrm{M}_U - \widehat{\mathrm{M}}_U)\,\mathrm{M}^k_X\right\|^2_{\mathbb{L}^2(\mathrm{t}_c)}\right].\end{aligned}$$

Analogously to estimating the density f, we obtain in Corollary 3.4 a risk representation with very similar structure as the corresponding risk representation of $\tilde{S}_k$ in (9) assuming the error density f^U to be known. The first and second term represents again the bias and variance as seen before. The last two summands, depending only on the additional measurements $\{U_j\}_{j\in[\![m]\!]}$, occur only due to the estimation of the unknown Mellin transformation M_U. Eventually, one observes directly, that the risk decomposition of $\hat{f}_k$ and $\widehat{S}_k^X$ are identical up to the density function $\mathrm{v} = \mathbb{1}$ and $\mathrm{v} = \mathrm{t}_c$, respectively. Hence, we subsequently study both cases simultaneously by considering a general $\mathbb{L}^2(\mathrm{v})$-risk for an arbitrary density function v, namely

$$\mathbb{E}\left[\left\|\widehat{\mathrm{M}}_X^k - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right]. \tag{26}$$

3.3. Oracle type inequalities and rates of convergences

We start by formalising assumptions needed to be satisfied in the subsequent parts.

Assumption B.I.
In addition to Assumption A.I let $\{U_j\}_{j\in[\![m]\!]}$ be i.i.d. copies of U, independently drawn from the sample $\{Y_i\}_{i\in[\![n]\!]}$ and let $f^U \in \mathbb{L}^1_+(\mathrm{x}^{4(c-1)})$. Moreover suppose that $\mathbb{1}_{[-k,k]} \in \mathbb{L}^2(\mathrm{v})$ for each $k \in \mathbb{R}_{>0}$. □

Corollary 3.5. *Under Assumption B.I, there exists a optimal tuning parameter $k_{\mathfrak{o}} \in \mathbb{R}_{>0}$, such that*

$$\begin{aligned}\mathbb{E}\left[\left\|\widehat{\mathrm{M}}_X^{k_{\mathfrak{o}}} - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right] &\leq 8\left(1 \vee \mathbb{E}[U_1^{2(c-1)}]\right)^2 \cdot \left(1 \vee \mathbb{E}[X_1^{2(c-1)}]\right) \\ &\quad\times \inf_{k\in\mathbb{R}_{>0}}\left\{\left\|\mathrm{M}_X \mathbb{1}_{[-k,k]^C}\right\|^2_{\mathbb{L}^2(\mathrm{v})} + \frac{1}{n}\left[\left\|\mathbb{1}_{[-k,k]}\mathrm{M}_U^{\dagger}\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right]\right\} \\ &\quad + 8\big(1 \vee \mathfrak{C}\,\mathbb{E}[U_1^{4(c-1)}]\big) \cdot \left\|\mathrm{M}_X\left(1 \vee |\,\mathrm{M}_U|^2 m\right)^{-1/2}\right\|^2_{\mathbb{L}^2(\mathrm{v})}. \end{aligned} \tag{27}$$

Proof of Corollary 3.5. Under Assumption B.I we start with the general $\mathbb{L}^2(\mathrm{v})$-risk representation for $k \in \mathbb{R}_{>0}$, given by

$$\begin{aligned}\mathbb{E}\left[\left\|\widehat{\mathrm{M}}_X^k - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right] &= \left\|\mathrm{M}_X \mathbb{1}_{[-k,k]^C}\right\|^2_{\mathbb{L}^2(\mathrm{v})} \\ &\quad + \frac{1}{n}\mathbb{E}\left[\left\|\widehat{\mathrm{M}}_U^{\dagger}\mathbb{1}_{\mathfrak{M}}\mathbb{V}_Y\mathbb{1}_{[-k,k]}\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right] \\ &\quad + \mathbb{E}\left[\left\|\mathrm{M}_X^k\,\mathbb{1}_{\mathfrak{M}^C}\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right] + \mathbb{E}\left[\left\|\widehat{\mathrm{M}}_U^{\dagger}\mathbb{1}_{\mathfrak{M}}(\mathrm{M}_U - \widehat{\mathrm{M}}_U)\,\mathrm{M}_X^k\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right],\end{aligned}$$

where the proof follows line by line of the proof of Lemma 3.2. We upper bound the last three summands of the last display with the help of Lemma B.1 in Appendix B and derive for each $k \in \mathbb{R}_{>0}$

$$\mathbb{E}\left[\left\|\widehat{\mathrm{M}}_X^k - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right] \leq \left\|\mathrm{M}_X \mathbb{1}_{[-k,k]^C}\right\|^2_{\mathbb{L}^2(\mathrm{v})}$$

$$+ 4\left(1 \vee \mathbb{E}[U_1^{2(c-1)}]\right)^2 \cdot \mathbb{E}[X_1^{2(c-1)}] \cdot \left\| \mathrm{M}_U^{\dagger}\, \mathbb{1}_{[-k,k]} \right\|^2_{\mathbb{L}^2(\mathrm{v})} \cdot n^{-1}$$
$$+ 4\big(1 \vee \mathbb{E}[U_1^{2(c-1)}]\big) \cdot \left\| \mathrm{M}_X^k \left(1 \vee |\,\mathrm{M}_U\,|^2 (m \wedge n)\right)^{-1/2} \right\|^2_{\mathbb{L}^2(\mathrm{v})}$$
$$+ 4\big(1 \vee \mathfrak{C}\,\mathbb{E}[U_1^{4(c-1)}]\big) \cdot \left\| \mathrm{M}_X^k \left(1 \vee |\,\mathrm{M}_U\,|^2 m\right)^{-1/2} \right\|^2_{\mathbb{L}^2(\mathrm{v})}.$$

Further, elementary computations show that

$$\left\| \mathrm{M}_X^k \left(1 \vee |\,\mathrm{M}_U\,|^2 (m \wedge n)\right)^{-1/2} \right\|^2_{\mathbb{L}^2(\mathrm{v})} \leq \left\| \mathrm{M}_X^k \left(1 \vee |\,\mathrm{M}_U\,|^2 m\right)^{-1/2} \right\|^2_{\mathbb{L}^2(\mathrm{v})} + \mathbb{E}[X_1^{2(c-1)}] \left\| \mathrm{M}_U^{\dagger}\, \mathbb{1}_{[-k,k]} \right\|^2_{\mathbb{L}^2(\mathrm{v})} \cdot n^{-1},$$

such that we finally obtain

$$\mathbb{E}\left[\left\| \widehat{\mathrm{M}}_X^k - \mathrm{M}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} \right] \leq \left\| \mathrm{M}_X\, \mathbb{1}_{[-k,k]^C} \right\|^2_{\mathbb{L}^2(\mathrm{v})}$$
$$+ 8\left(1 \vee \mathbb{E}[U_1^{2(c-1)}]\right)^2 \cdot \mathbb{E}[X_1^{2(c-1)}] \cdot \left\| \mathrm{M}_U^{\dagger}\, \mathbb{1}_{[-k,k]} \right\|^2_{\mathbb{L}^2(\mathrm{v})} \cdot n^{-1}$$
$$+ 8\big(1 \vee \mathfrak{C}\,\mathbb{E}[U_1^{4(c-1)}]\big) \left\| \mathrm{M}_X \left(1 \vee |\,\mathrm{M}_U\,|^2 m\right)^{-1/2} \right\|^2_{\mathbb{L}^2(\mathrm{v})}.$$

The claim follows from the last decomposition, where the first two terms in the upper bound depend on $k \in \mathbb{R}_{>0}$ only, and observing that the first term decreases while the second one increases for an increasing k. ☐

Observe that in the oracle-type inequality (27) the upper bound consists of two summands. The first summand is identical up to the constants to the upper bound in (11) and we refer to its discussion in Subsection 2.5. Again, $k_{\mathfrak{o}}$ represents an oracle choice, as it depends on the unknown Mellin transformation M_X. We should emphasise that the first summand only depends on the sample size n, while the second summand only the size m of the additional sample. Therefore the second represents the additional cost of estimating M_U as well. In case of additive convolution on the circle and the real line [21] and [25] derive respectively a oracle-type inequality for integrated mean squared error, i.e. the global $\mathbb{L}^2$-risk, with an upper bound consisting also of two terms, each one depending on one sample size only.
Similar to Subsection 2.5 we assume in the following brief discussion an unknown density $f \in \mathbb{W}^s(L)$, where the regularity s is specified below. Regarding the Mellin transformation M_U, we subsequently assume again its *ordinary smoothness* ($\mathfrak{o.s.}$) or *super smoothness* ($\mathfrak{s.s.}$). In order to discuss the convergences rates for the $\mathbb{L}^2(\mathrm{v})$-risk under these regularity assumptions, we restrict ourselves again to the choice $\mathrm{v} := \mathrm{t}_c^{-a}$ for $a \in \mathbb{R}$, observing that $a = 0$ corresponds to the global risk for estimating the density f and $a = -1$ corresponds to estimating the survival function S^X of X as discussed before.

Corollary 3.6. *Under Assumption B.I, let $f \in \mathbb{W}^s(L)$ for some fixed $L \in \mathbb{R}_{>0}$.*

1. *If* M_U *satisfies eq.* ($\mathfrak{o.s.}$), *choose* $k_{\mathfrak{o}} := n^{\frac{1}{2\gamma+2s+1}}$.
   Then, for any $a \in (-1/2-\gamma, s)$,

$$\mathbb{E}\left[\left\|\widehat{\mathrm{M}}_X^{k_{\mathfrak{o}}} - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{t}_c^{-a})}\right] \leq \mathfrak{C}_1 \cdot \left(n^{-\frac{2(s-a)}{2\gamma+2s+1}} + m^{-(\frac{s-a}{\gamma}\wedge 1)}\right),$$

   where $\mathfrak{C}_1 = \mathfrak{C}_1(c, a, s, \mathfrak{c}_l, \mathbb{E}[X_1^{2(c-1)}], \mathbb{E}[U_1^{4(c-1)}]) \in \mathbb{R}_{>0}$.
2. *If* M_U *satisfies eq.* ($\mathfrak{s.s.}$), *choose* $k_{\mathfrak{o}} := (\log n)^{\frac{1}{2\gamma}}$. *Then, for any* $a < s$

$$\mathbb{E}\left[\left\|\widehat{\mathrm{M}}_X^{k_{\mathfrak{o}}} - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{t}_c^{-a})}\right] \leq \mathfrak{C}_2 \cdot \left((\log n)^{-\frac{s-a}{\gamma}} + (\log m)^{-\frac{s-a}{\gamma}}\right),$$

   where $\mathfrak{C}_2 = \mathfrak{C}_2(c, a, s, \mathfrak{c}_l, \mathbb{E}[X_1^{2(c-1)}], \mathbb{E}[U_1^{4(c-1)}]) \in \mathbb{R}_{>0}$.

Proof of Corollary 3.6. Starting with Corollary 3.5 the upper bound for the first term is given in Corollary 2.2. Regarding the second summand, assume first that M_U is ordinary smooth, i.e. eq. ($\mathfrak{o.s.}$) holds, and $a+\gamma+\frac{1}{2} > 0$. Since M_X belongs to the Mellin-Sobolev ellipsoid $\mathbb{W}^s(L)$ with $a < s$

$$\begin{aligned}
&\left\|\mathrm{M}_X\left(1 \vee |\mathrm{M}_U|^2 m\right)^{-1/2}\right\|^2_{\mathbb{L}^2(\mathrm{t}_c^{-a})} \leq \mathfrak{c}_l L^2 \left\|\mathrm{t}_2^s \mathrm{t}_c^{-a}\left(1 \vee \mathrm{t}_2^{-\gamma} m\right)^{-1}\right\|_{\mathbb{L}^\infty} \\
&\leq \mathfrak{C}(c, a, s, \mathfrak{c}_l, L) \cdot \left\|\mathrm{t}_2^{s-a}\left(1 \vee \mathrm{t}_2^{-\gamma} m\right)^{-1}(\mathbb{1}_{[-m^{\frac{1}{2\gamma}}, m^{\frac{1}{2\gamma}}]} + \mathbb{1}_{[-m^{\frac{1}{2\gamma}}, m^{\frac{1}{2\gamma}}]^C})\right\|_{\mathbb{L}^\infty} \\
&\leq \mathfrak{C}(c, a, s, \mathfrak{c}_l, L) \cdot \max\left\{m^{\frac{(a-s+\gamma)_+ - \gamma}{\gamma}}, m^{-\frac{s-a}{\gamma}}\right\}, \qquad (28)
\end{aligned}$$

which shows the first claim. Similarly, if M_U is super smooth, i.e. eq. ($\mathfrak{s.s.}$) holds, we have for any $a < s$ setting $I := [-(\log m)^{\frac{1}{2\gamma}}, (\log m)^{\frac{1}{2\gamma}}]$

$$\begin{aligned}
&\left\|\mathrm{M}_X\left(1 \vee |\mathrm{M}_U|^2 m\right)^{-1/2}\right\|^2_{\mathbb{L}^2(\mathrm{t}_c^{-a})} \leq \mathfrak{c}_l L^2 \left\|\mathrm{t}_2^s \mathrm{t}_c^{-a}\left(1 \vee \exp(-2\mathfrak{c}_l|\cdot|^{2\gamma}) m\right)^{-1}\right\|_{\mathbb{L}^\infty} \\
&\leq \mathfrak{C}(c, a, s, \mathfrak{c}_l, L) \cdot \left\|\mathrm{t}_2^{s-a}\left(1 \vee \exp(-\mathrm{t}_2^{-\gamma}) m\right)^{-1}(\mathbb{1}_I + \mathbb{1}_{I^C})\right\|_{\mathbb{L}^\infty} \\
&\leq \mathfrak{C}(c, a, s, \mathfrak{c}_l, L) \cdot \max\left\{(\log m)^{-\frac{s-a}{\gamma}}, (\log m)^{-\frac{s-a}{\gamma}}\right\}, \qquad (29)
\end{aligned}$$

which shows the second claim and completes the proof. □

Remark 3.1. Let us briefly discuss the choices of $a = 0$ and $a = -1$, which provide upper bounds for the global estimation risk for the density f and the survival function S^X, respectively:

1. Density estimation: Choosing $a = 0$ for all $f \in \mathbb{W}^s(L)$ leads to

$$\mathbb{E}\left[\left\|\hat{f}_{\hat{k}} - f\right\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})}\right] \lesssim n^{-\frac{2s}{2\gamma+2s+1}} + m^{-(\frac{s}{\gamma}\wedge 1)}$$

   if M_U is ordinary smooth, and

$$\mathbb{E}\left[\left\|\hat{f}_{\hat{k}} - f\right\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})}\right] \lesssim (\log n)^{-\frac{s}{\gamma}} + (\log m)^{-\frac{s}{\gamma}}$$

if M_U is super smooth.

2. Survival function estimation: Choosing $a=-1$ for all $f \in \mathbb{W}^s(L)$ leads to

$$\mathbb{E}\left[\left\|\widehat{S}_{\hat{k}}^X - S\right\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-3})}\right] \lesssim n^{-\frac{2(s+1)}{2\gamma+2s+1}} + m^{-(\frac{s+1}{\gamma}\wedge 1)}$$

if M_U is ordinary smooth, and

$$\mathbb{E}\left[\left\|\widehat{S}_{\hat{k}}^X - S\right\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-3})}\right] \lesssim (\log n)^{-\frac{s+1}{\gamma}} + (\log m)^{-\frac{s+1}{\gamma}}$$

if M_U is super smooth.

There is no corresponding lower bound of the additional second term in the upper bound given in Corollary 3.6, which only depends on the second sample size m. Nevertheless, a similar expression occurs in circular additive deconvolution with unknown error density, where in [21] a matching lower bound is presented. Moreover, in additive deconvolution on the real line with unknown error distribution the additional second term is also present and [25] derives a matching lower bound in case $a=0$. This suggests that this is also the case in the multiplicative measurement model with continuous Mellin Transform, although a proof is missing yet. □

4. Data driven estimation

In this section we will provide a fully data-driven selection method for k based on the construction given in Subsection 2.6 but dismissing the knowledge of the error density f^U. A similar approach has been considered for additive deconvolution problems for instance in [16] and [21]. More precisely, we select

$$\hat{k} :\in \underset{k\in[\![k_n]\!]}{\arg\min}\left\{-\left\|\mathbb{1}_{[-k,k]}\widehat{\mathrm{M}}_Y\widehat{\mathrm{M}}_U^{\dagger}\mathbb{1}_{\mathfrak{M}}\right\|^2_{\mathbb{L}^2(\mathrm{v})} + 2\widehat{\sigma}_Y^2\,\mathrm{pen}_k^{\hat{\mathrm{v}}}\right\}$$
$$= \underset{k\in[\![k_n]\!]}{\arg\min}\left\{-\left\|\widehat{\mathrm{M}}_X^k\right\|^2_{\mathbb{L}^2(\mathrm{v})} + 2\widehat{\sigma}_Y^2\,\mathrm{pen}_k^{\hat{\mathrm{v}}}\right\},$$

where the penalisation term $\mathrm{pen}_k^{\hat{\mathrm{v}}}$ depends on a *random* density function $\hat{\mathrm{v}}$, which is defined and specified later as well as the choice of $k_n \in \mathbb{N}$. In contrast to the selection (18) in Subsection 2.6, here we have replaced $\mathrm{M}_U^{\dagger}$ by its empirical counterpart $\widehat{\mathrm{M}}_U^{\dagger}$. Our aim is to analyse the global $\mathbb{L}^2(\mathrm{v})$-risk again, namely

$$\mathbb{E}\left[\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right]. \tag{30}$$

Our upper bounds necessities also slightly stronger assumptions than Assumption B.I, which we formulate next.

Assumption B.II.
In addition to Assumptions A.II and B.I let $\inf_{t\in[-1,1]}\{|\,\mathrm{M}_U(t)|\} \in \mathbb{R}_{>0}$, $f^U \in \mathbb{L}^1_+(\mathrm{x}^{7(c-1)}) \cap \mathbb{L}^1_+(\mathrm{x}^{2(c-1)}|\log \mathrm{x}|^{\gamma})$ for some $\gamma \in \mathbb{R}_{>0}$, $f^Y \in \mathbb{L}^1_+(\mathrm{x}^{16(c-1)})$, and such that there are $\eta_Y, \eta_U, \eta_X \in \mathbb{R}_{\geq 1}$ satisfying

i) $\eta_Y \geq \max\left\{\|f^Y\|_{\mathbb{L}^\infty_+(\mathrm{x}^{2c-1})}, \|f^Y\|_{\mathbb{L}^1_+(\mathrm{x}^{16(c-1)})}\right\}$,
ii) $\eta_U \geq \max\left\{\|f^U\|^{1/7}_{\mathbb{L}^1_+(\mathrm{x}^{7c-1})}, \|f^U\|_{\mathbb{L}^1_+(\mathrm{x}^{2(c-1)}|\log \mathrm{x}|^{\gamma})}\right\}$,
iii) $\eta_X \geq \max\left\{\|f^X\|_{\mathbb{L}^1_+(\mathrm{x}^{2(c-1)})}, \|\mathrm{M}_X\|_{\mathbb{L}^2(\mathrm{v})}\right\}$.

We set $a_Y \in \mathbb{R}_{>0}$ and $k_Y \in \mathbb{R}_{\geq 1}$ as in Assumption A.II. □

Motivated by the key argument in (20) in case of a known error density f^U, the next lemma provides an error bound when f^U is unknown. Its proof can be found in Appendix C.

Lemma 4.1. *Consider an arbitrary event $\mho$ with complement $\mho^C$, and denote by $\mathfrak{A}^C$ the complement of the event $\mathfrak{A} := \{\sigma_Y^2 \leq 2\widehat{\sigma}_Y^2\}$. Given $k_n \in \mathbb{R}_{\geq 1}$ and*

$$\hat{k} :\in \underset{k\in[\![k_n]\!]}{\arg\min}\left\{-\left\|\mathbb{1}_{[-k,k]}\widehat{\mathrm{M}}_Y\widehat{\mathrm{M}}_U^{\dagger}\mathbb{1}_{\mathfrak{M}}\right\|^2_{\mathbb{L}^2(\mathrm{v})} + 2\widehat{\sigma}_Y^2\,\mathrm{pen}_k^{\widehat{\mathrm{v}}}\right\} \tag{31}$$

for any $k_\circ \in [\![k_n]\!]$ we have

$$\begin{aligned}\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})} &\leq 15\left\|\mathbb{1}_{[-k_\circ,k_\circ]}(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\widehat{\mathrm{M}}_U^{\dagger}\mathbb{1}_{\mathfrak{M}}\right\|^2_{\mathbb{L}^2(\mathrm{v})}\\ &+ 15\left\|\mathbb{1}_{[-k_\circ,k_\circ]^C}\,\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})} + 24\widehat{\sigma}_Y^2\,\mathrm{pen}_{k_\circ}^{\widehat{\mathrm{v}}}\,\mathbb{1}_{\mho} + 6\left\|\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})}\mathbb{1}_{\mho^c}\\ &+ 15\left\|(\mathrm{M}_U\,\widehat{\mathrm{M}}_U^{\dagger}\mathbb{1}_{\mathfrak{M}} - \mathbb{1})\,\mathrm{M}_X^{k_n}\right\|^2_{\mathbb{L}^2(\mathrm{v})}\\ &+ 12\max_{k\in[\![k_n]\!]}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\widehat{\mathrm{M}}_U^{\dagger}\mathbb{1}_{\mathfrak{M}}\right\|^2_{\mathbb{L}^2(\mathrm{v})} - \tfrac{\sigma_Y^2}{4}\,\mathrm{pen}_k^{\widehat{\mathrm{v}}}\right)_+\right\}\\ &+ 3\left\|\mathbb{1}_{[-k_n,k_n]}(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\widehat{\mathrm{M}}_U^{\dagger}\mathbb{1}_{\mathfrak{M}}\right\|^2_{\mathbb{L}^2(\mathrm{v})}(\mathbb{1}_{\mho^c} + \mathbb{1}_{\mathfrak{A}^c}).\end{aligned} \tag{32}$$

In the sequel, we aim to apply the expectation on both sides of (32) in order to derive an upper bound for the risk. Therefore, we need to control amongst others the expectation of

$$\max_{k\in[\![k_n]\!]}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\widehat{\mathrm{M}}_U^{\dagger}\mathbb{1}_{\mathfrak{M}}\right\|^2_{\mathbb{L}^2(\mathrm{v})} - \tfrac{\sigma_Y^2}{4}\,\mathrm{pen}_k^{\widehat{\mathrm{v}}}\right)_+\right\}. \tag{33}$$

In Subsection 2.6 a similar term was controlled by introducing the density function $\mathrm{v}_U := |\,\mathrm{M}_U^{\dagger}\,|^2\mathrm{v}$ and providing a concentration inequality in Proposition 2.3. In contrast, we use in the sequel its empirical counterpart, the random density function $\widehat{\mathrm{v}} := |\widehat{\mathrm{M}}_U^{\dagger}\mathbb{1}_{\mathfrak{M}}|^2\mathrm{v}$, which depends on the sample $\{U_j\}_{j\in[\![m]\!]}$ only. Then (33) reads as

$$\max_{k\in[\![k_n]\!]}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y-\mathrm{M}_Y)\widehat{\mathrm{M}}_U^{\dagger}\mathbb{1}_{\mathfrak{M}}\right\|^2_{\mathbb{L}^2(\mathrm{v})}-\frac{\sigma_Y^2}{4}\operatorname{pen}_k^{\hat{\mathrm{v}}}\right)_+\right\}$$
$$=\max_{k\in[\![k_n]\!]}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y-\mathrm{M}_Y)\right\|^2_{\mathbb{L}^2(\hat{\mathrm{v}})}-\frac{\sigma_Y^2}{4}\operatorname{pen}_k^{\hat{\mathrm{v}}}\right)_+\right\}.$$

The next proposition provides a concentration inequality for the expectation of the quantity in the last display and its proof can be found in Appendix C.

Proposition 4.2 (Concentration inequality)**.** *Under Assumption B.II for* v *and* $\hat{\mathrm{v}}:=|\widehat{\mathrm{M}}_U^{\dagger}|^2\mathrm{v}$ *define* Δ_k^{v} *and* δ_k^{v} *as well as* $\Delta_k^{\hat{\mathrm{v}}}$ *and* $\delta_k^{\hat{\mathrm{v}}}$ *as in* (17)*. We consider* k_n*,* $n\in\mathbb{N}$*, defined by*

$$k_n:=\max\{k\in[\![n]\!]:k\Delta_k^{\mathrm{v}}\le n\Delta_1^{\mathrm{v}}\}. \tag{34}$$

We then have

$$\mathbb{E}\left[\max_{k\in[\![k_n]\!]}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y-\mathrm{M}_Y)\right\|^2_{\mathbb{L}^2(\hat{\mathrm{v}})}-12\sigma_Y^2\Delta_k^{\hat{\mathrm{v}}}\delta_k^{\hat{\mathrm{v}}}kn^{-1}\right)_+\right\}\right]$$
$$\le\mathfrak{C}\cdot\eta_Y(1\vee\eta_U^2k_Y\Delta_{k_Y}^{\mathrm{v}})\cdot n^{-1}.$$

Following the argumentation in Subsection 2.6, we choose $k_n\in[\![n]\!]$ as in (34). Consequently, defining $\operatorname{pen}_k^{\hat{\mathrm{v}}}$ appropriately, namely for each $k\in[\![k_n]\!]$ by

$$\operatorname{pen}_k^{\hat{\mathrm{v}}}:=24\Delta_k^{\hat{\mathrm{v}}}\delta_k^{\hat{\mathrm{v}}}kn^{-1}, \tag{35}$$

Proposition 4.2 allows us to upper bound the expectation of eq. (33). Thus, the data-driven dimension parameter is specified as follows,

$$\hat{k}:\in\underset{k\in[\![k_n]\!]}{\arg\min}\left\{-\left\|\widehat{\mathrm{M}}_X^k\right\|^2_{\mathbb{L}^2(\hat{\mathrm{v}})}+48\widehat{\sigma}_Y^2\Delta_k^{\hat{\mathrm{v}}}\delta_k^{\hat{\mathrm{v}}}k\right\},$$

where $\widehat{\sigma}_Y^2$ is again the unbiased estimator of σ_Y^2 given in Subsection 2.6. For any $k_\circ\in[\![k_n]\!]$ we intend to apply Lemma 4.1 with the random set

$$\mho:=\mho_{k_\circ}:=\left\{\sup_{t\in[-k_\circ,k_\circ]}\left|\widehat{\mathrm{M}}_U(t)\,\mathrm{M}_U(t)-1\right|\le\frac{1}{3}\right\}$$
$$\subseteq\left\{\sup_{t\in[-k_\circ,k_\circ]}\left|\widehat{\mathrm{M}}_U^{\dagger}(t)\,\mathrm{M}_U(t)\right|^2\le\frac{9}{4}\right\} \tag{36}$$

where its complement evidently satisfies

$$\mho_{k_\circ}^C=\left\{\exists t\in[-k_\circ,k_\circ]:\left|\widehat{\mathrm{M}}_U(t)-\mathrm{M}_U(t)\right|>\frac{1}{3}|\,\mathrm{M}_U(t)|\right\}.$$

The following lemma provides a first upper bound of the risk, which follows directly by applying the expectation on both sides of Lemma 4.1 as well as the concentration inequality in Proposition 4.2. The proof with all details can be found in Appendix C.

Lemma 4.3. *Under Assumption B.II for any $k_\circ \in [\![k_n]\!]$ we have*

$$\begin{aligned}\mathbb{E}\left[\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right] &\\ \leq \mathfrak{C}\cdot(\eta_U^4\eta_X+\sigma_Y^2)&\min_{k\in[\![k_n]\!]}\left\{\left\|\mathbb{1}_{[-k,k]^C}\,\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})}+\Delta_k^{\mathrm{v}_U}\delta_k^{\mathrm{v}_U}kn^{-1}\right\}\\ &+\mathfrak{C}\cdot\eta_U^4\left\|\mathrm{M}_X\left(1\vee|\mathrm{M}_U|^2m\right)^{-1/2}\right\|^2_{\mathbb{L}^2(\mathrm{v})}+\mathfrak{C}\cdot\eta_Y(1\vee\eta_U^2k_Y\Delta_{k_Y}^{\mathrm{v}})\cdot n^{-1}\\ &+15\cdot\left\|\mathbb{1}_{[-k_\circ,k_\circ]^C}\,\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})}+6\cdot(\sigma_Y^2\Delta_1^{\mathrm{v}}+\eta_X)\mathbb{P}(\mho_{k_\circ}^C).\end{aligned}\tag{37}$$

It remains to bound the probability of the event $\mho_{k_\circ}^C$, which turns out to be rather involved. The proof of the upper bound is again based on an inequality due to [28] which in the form of Lemma C.1 in the Appendix C for example is stated by [7] in equation (5.13) in Corollary 2. However, its application necessitates to bound the expectation of the supremum of a normalised Mellin function process which we establish in the next lemma. Its proof follows along the lines of the proof of Theorem 4.1 in [26] where a similar result for a normalised characteristic function process is shown. The proof of Proposition 4.4 is also postponed to the Appendix C.

Proposition 4.4. *Let $\{Z_j\}_{j\in\mathbb{N}}$ be a family of i.i.d $\mathbb{R}_{>0}$-valued random variables and assume there exists a constant $\eta\in\mathbb{R}_{\geq 1}$, such that $\eta \geq (\mathbb{E}[Z_1^{2\beta}])^{1/2}$ and $\eta\geq\mathbb{E}[Z_1^{2\beta}|\log(Z_1)|^\gamma]$ for some $\beta\in\mathbb{R}$ and $\gamma\in\mathbb{R}_{>0}$. Define the normalised Mellin function process by*

$$c_m(t) := \frac{1}{\sqrt{m}}\sum_{j\in[\![m]\!]}\left\{Z_j^{\beta+\iota 2\pi t}-\mathbb{E}\left[Z_j^{\beta+\iota 2\pi t}\right]\right\},\quad \forall t\in\mathbb{R},\tag{38}$$

and for $\rho\in\mathbb{R}_{>0}$ the density function $\bar{\mathrm{w}}:\mathbb{R}\to(0,1]$ by

$$\bar{\mathrm{w}}(t) := (\log(e+|t|))^{-\frac{1}{2}-\rho},\quad \forall t\in\mathbb{R}.\tag{39}$$

Then there exists a constant $\mathfrak{C}(\eta,\rho)\in\mathbb{R}_{\geq 1}$ only depending on η and ρ, such that

$$\sup_{m\in\mathbb{N}}\left\{\mathbb{E}\left[\|c_m\|_{\mathbb{L}^\infty(\bar{\mathrm{w}})}\right]\right\}\leq\mathfrak{C}(\eta,\rho).$$

Now, we are in the position to state an upper bound of the probability of the event $\mho_{k_\circ}^C$, which is proven in the Appendix C.

Proposition 4.5. *Let Assumption B.II be satisfied and let $\rho\in\mathbb{R}_{>0}$ be arbitrary but fixed. Consider the density function $\bar{\mathrm{w}}:\mathbb{R}\to(0,1]$ given in (39) and let $\mathfrak{C}(\eta_U,\rho)\in\mathbb{R}_{\geq1}$ be given by Proposition 4.4 (with $\beta = c-1$). Given the universal numerical constant $\mathfrak{C}_{\mathrm{tal}}\in\mathbb{R}_{>0}$ determined by Talagrand's inequality in Lemma C.1 we set*

$$\tau_m := \tau_m(\eta_U,\rho) := 2\eta_U\mathfrak{C}_{\mathrm{tal}}^{-1/2}(\log m)^{1/2}+2\mathfrak{C}(\eta_U,\rho)\quad\forall m\in\mathbb{N}\quad\textit{and}$$

$$m_\circ := m_\circ(\mathrm{M}_U, \eta_U, \rho) \\ := \min\left\{ m \in \mathbb{N}_{\geq 3} : \inf_{t \in [-1,1]} \bar{\mathrm{w}}(t) |\,\mathrm{M}_U(t)| \geq 6\tau_m m^{-1/2} \wedge \mathfrak{C}_{\mathrm{tal}} \eta_U^2 m \geq \log m \right\}. \tag{40}$$

For $m \in [\![m_\circ]\!]$ let $k_m \in \mathbb{R}_{>0}$ be arbitrary while for $m \in \mathbb{N}_{\geq m_\circ}$ we set

$$k_m := \sup\left\{ k \in \mathbb{N} : \inf_{t \in [-k,k]} \bar{\mathrm{w}}(t) |\,\mathrm{M}_U(t)| \geq 6\tau_m m^{-1/2} \right\}, \tag{41}$$

where the defining set is not empty for all $m \in \mathbb{N}_{\geq m_\circ}$. For any $m \in \mathbb{N}$ consider the event

$$\mho_{k_m}^C := \left\{ \exists t \in [-k_m, k_m] : \left| \widehat{\mathrm{M}}_U(t) - \mathrm{M}_U(t) \right| > \frac{1}{3} |\,\mathrm{M}_U(t)| \right\}.$$

then there is an universal numerical constant $\mathfrak{C} := 3 + 11\mathfrak{C}_{\mathrm{tal}}^{-3} \in \mathbb{R}_{\geq 1}$ such that we have

$$\mathbb{P}(\mho_{k_m}^C) \leq (m_\circ^2 \vee \mathfrak{C}\eta_U)\, m^{-2}, \quad \forall m \in \mathbb{N}.$$

We can now formulate our main result.

Theorem 4.6. *Let Assumption B.II be satisfied and let $\rho \in \mathbb{R}_{>0}$ be arbitrary but fixed. Consider $m_\circ$ as in (40) and for each $m, n \in \mathbb{N}$, k_n as in (34) and k_m as in (41). Then for all $m, n \in \mathbb{N}$ we have*

$$\begin{aligned} &\mathbb{E}\left[\left\| \widehat{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \right] \\ &\quad \leq \mathfrak{C} \cdot (\eta_U^4 \eta_X + \sigma_Y^2) \min_{k \in [\![k_n]\!]} \left\{ \left\| \mathbb{1}_{[-k,k]^C} \mathrm{M}_X \right\|_{\mathbb{L}^2(\mathrm{v})}^2 + \Delta_k^{\mathrm{v}_U} \delta_k^{\mathrm{v}_U} k n^{-1} \right\} \\ &\qquad + \mathfrak{C} \cdot \eta_U^4 \left\| \mathrm{M}_X \left(1 \vee |\,\mathrm{M}_U|^2 m \right)^{-1/2} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 + \mathfrak{C} \cdot \eta_Y (1 \vee \eta_U^2 k_Y \Delta_{k_Y}^{\mathrm{v}}) \cdot n^{-1} \\ &\qquad + 15 \cdot \left\| \mathbb{1}_{[-k_m,k_m]^C} \mathrm{M}_X \right\|_{\mathbb{L}^2(\mathrm{v})}^2 + \mathfrak{C} \cdot (\sigma_Y^2 \Delta_1^{\mathrm{v}} + \eta_X)(m_\circ^2 \vee \eta_U)\, m^{-1}. \end{aligned}$$

Proof of Theorem 4.6. By combining Lemma 4.3 with $k_\circ := k_n \wedge k_m$, the elementary bound

$$\begin{aligned} \left\| \mathbb{1}_{[-k_\circ,k_\circ]^C} \mathrm{M}_X \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \leq \min_{k \in [\![k_n]\!]} &\left\{ \left\| \mathbb{1}_{[-k,k]^C} \mathrm{M}_X \right\|_{\mathbb{L}^2(\mathrm{v})}^2 + \Delta_k^{\mathrm{v}_U} \delta_k^{\mathrm{v}_U} k n^{-1} \right\} \\ &+ \left\| \mathbb{1}_{[-k_m,k_m]^C} \mathrm{M}_X \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \end{aligned}$$

and $\mathbb{P}(\mho_{k_\circ}^C) \leq \mathbb{P}(\mho_{k_m}^C) \leq (m_\circ^2 \vee \mathfrak{C}\eta_U)\, m^{-2}$ we immediately obtain the claim, which completes the proof. ☐

Remark 4.1. Let us briefly compare the upper bound for the risk of the data-driven estimator with known and unknown error distribution given in (19) and in

Theorem 4.6, respectively. Up to the constants estimating the error distribution leads to the three additional terms

$$\left\| \mathrm{M}_X \left(1 \vee |\,\mathrm{M}_U\,|^2 m\right)^{-1/2} \right\|_{\mathbb{L}^2(\mathrm{v})}^2, \; m^{-1}, \text{ and } \left\| \mathbb{1}_{[-k_m,k_m]^C}\, \mathrm{M}_X \right\|_{\mathbb{L}^2(\mathrm{v})}^2. \tag{42}$$

Evidently, they depend all on the sample size m only, and the second is negligible with respect to the first term. Moreover, the first term in (42) is already present in the upper risk bound (27) of the oracle estimator when estimating the error density too. Thus the third term in (42) characterises the prize we pay for selecting the dimension parameter fully data-driven while estimating the error density. Comparing the first and third term in (42) it is not obvious to us which one is the leading term. If there exists a constant $\mathfrak{C} \in \mathbb{R}_{>1}$ such that

$$[-k_m, k_m]^C \subset \{\bar{\mathrm{w}}^2 |\,\mathrm{M}_U\,|^2 m (\log m)^{-1} < \mathfrak{C}\}, \quad \forall m \in \mathbb{N}, \tag{43}$$

then both, the first and third term in (42), are bounded up to the constant $\mathfrak{C}$ by

$$\left\| \mathrm{M}_X \left(1 \vee \bar{\mathrm{w}}^2 |\,\mathrm{M}_U\,|^2 m / \log m\right)^{-1/2} \right\|_{\mathbb{L}^2(\mathrm{v})}^2. \tag{44}$$

Note that the additional condition (43) is satisfied whenever $|\,\mathrm{M}_U\,|^2$ is monotonically decreasing. However, the term in (44) might over estimated both, the first and third term in (42). Take for example the situation $\mathfrak{o.s.}$ in Corollary 4.7 below. In case $(s-a) > \gamma$ all three terms in (42) are of order m^{-1} while the term in (44), by similar computations as in the proof of Corollary 3.6, is of order $m^{-1}(\log m)^{2(1+\rho)}$. In contrast in the situation $\mathfrak{s.s.}$ in Corollary 4.7 below, the first and third term in (42) as well as the term in (44) are of order $(\log m)^{-\frac{s-a}{\gamma}}$. □

Continuing the brief discussion in Subsection 3.3 we assume an unknown density $f \in \mathbb{W}^s(L)$, where the regularity s is specified below. Regarding the Mellin transformation M_U, we subsequently assume again its *ordinary smoothness* ($\mathfrak{o.s.}$) or *super smoothness* ($\mathfrak{s.s.}$). In order to discuss the convergences rates for the $\mathbb{L}^2(\mathrm{v})$-risk under these regularity assumptions, we restrict ourselves again to the choice $\mathrm{v} := \mathrm{t}_c^{-a}$ for $a \in \mathbb{R}$, observing that $a = 0$ corresponds to the global risk for estimating the density f and $a = -1$ corresponds to estimating the survival function S^X of X as discussed before.

Corollary 4.7. *Under the assumptions and notations of Theorem 4.6, we have the following rates of convergences.*

1. *If* M_U *satisfies eq.* ($\mathfrak{o.s.}$), *then,*

$$\mathbb{E}\left[\left\| \widehat{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X \right\|_{\mathbb{L}^2(\mathrm{t}_c^{-a})}^2\right] \leq \mathfrak{C}_1 \left(n^{-\frac{2(s-a)}{2\gamma+2s+1}} + \left(\frac{(\log m)^{2(1+\rho)}}{m}\right)^{\left(\frac{s-a}{\gamma}\right)} \vee m^{-1} \right),$$

for any $a \in (-1/2-\gamma, s)$, *where* $\mathfrak{C}_1 = \mathfrak{C}_1(c, a, s, L, \sigma_Y^2, \eta_X, \eta_U, \eta_Y, k_Y)$.

2. *If* M_U *satisfies eq.* ($\mathfrak{s.s.}$), *then,*

$$\mathbb{E}\left[\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{t}_c^{-a})}\right] \leq \mathfrak{C}_2 \cdot \left((\log n)^{-\frac{s-a}{\gamma}} + (\log m)^{-\frac{s-a}{\gamma}}\right),$$

for any $a < s$, *where* $\mathfrak{C}_2 = \mathfrak{C}_2(c, a, s, L, \sigma_Y^2, \eta_X, \eta_U, \eta_Y, k_Y)$.

Proof of Corollary 4.7. Starting with Theorem 4.6 the upper bounds for the first summand is given in the proof of Corollary 2.5 and for the second in Corollary 3.6. It remains to upper bound the fourth summand, by using eq. (14)

$$\left\|\mathbb{1}_{[-k_m,k_m]^C}\,\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})} \leq \mathfrak{C}(c, a, s, L) k_m^{-2(s-a)}, \tag{45}$$

since M_X belongs to the Mellin-Sobolev ellipsoid $\mathbb{W}^s(L)$ with $a < s$. The value of $k_m \in \mathbb{N}$ is determined by eq. (41) in Proposition 4.5. We note that there are finite constants $\mathfrak{C}_1 = \mathfrak{C}_1(\eta_U, \rho) \in \mathbb{R}_{>0}$ and $\mathfrak{C}_2 = \mathfrak{C}_2(\rho) \in \mathbb{R}_{>0}$, such that $\tau_m \leq \mathfrak{C}_1 (\log m)^{\frac{1}{2}}$ for all $m \in \mathbb{N}_{\geq m_\circ}$ and

$$\inf_{t\in[-k,k]} |\bar{\mathrm{w}}(t)| (\log k)^{\frac{1}{2}+\rho} \geq \mathfrak{C}_2,$$

for all $k \in \mathbb{R}_{\geq 3}$. Assume that M_U is ordinary smooth, i.e. eq. ($\mathfrak{o.s.}$) is satisfied, then there exists a finite constant $\mathfrak{C}_3 = \mathfrak{C}_3(\mathfrak{c}_l, \gamma) \in \mathbb{R}_{>0}$, such that

$$\inf_{t\in[-k,k]} |\,\mathrm{M}_U(t)| k^{\gamma} \geq \mathfrak{C}_3,$$

for all $k \in \mathbb{R}_{\geq 1}$. Consequently, there exists a constant $\mathfrak{C}_4 = \mathfrak{C}_4(\mathfrak{c}_l, \gamma, \eta_U, \rho) \in \mathbb{R}_{>0}$ small enough, such that for $k = \left\lfloor \mathfrak{C}_4 m^{\frac{1}{2\gamma}} (\log m)^{\frac{-1-\rho}{\gamma}} \right\rfloor$

$$\begin{aligned}\inf_{t\in[-k,k]} |\bar{\mathrm{w}}(t)||\,\mathrm{M}_U(t)| &\geq \mathfrak{C}_2 \mathfrak{C}_3 k^{-\gamma} (\log k)^{-\frac{1}{2}-\rho} \\ &\geq 6 m^{-\frac{1}{2}} \mathfrak{C}_1 (\log m)^{\frac{1}{2}} \geq 6 \tau_m m^{-\frac{1}{2}}.\end{aligned}$$

Therefore, exploiting eq. (41), i.e. $k_m \geq k$ and eq. (45), we obtain

$$\left\|\mathbb{1}_{[-k_m,k_m]^C}\,\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})} \leq \mathfrak{C}_5(c, a, s, L, \mathfrak{c}_l, \gamma, \eta_U, \rho) \left(\frac{(\log m)^{2+2\rho}}{m}\right)^{\frac{s-a}{\gamma}}.$$

Considering Theorem 4.6 and combining the last upper bound and the upper bounds in eqs. (23) and (28) in the proofs of Corollaries 2.5 and 3.6, respectively, we obtain the first claim. Secondly, assume that M_U is supper smooth, i.e. eq. ($\mathfrak{s.s.}$) is satisfied, then

$$\inf_{t\in[-k,k]} |\,\mathrm{M}_U(t)| \exp(\mathfrak{c}_l k^{2\gamma}) \geq 1,$$

for all $k \in \mathbb{R}_{\geq 1}$. Consequently, there exists a constant $\mathfrak{C}_4 = \mathfrak{C}_4(\mathfrak{c}_l, \gamma, \eta_U, \rho) \in \mathbb{R}_{>0}$ small enough, such that for $k = \left\lfloor \left(\frac{1}{2\mathfrak{c}_l} \log \frac{\mathfrak{C}_4 \cdot m}{(\log m)^{2+2\rho}}\right)^{\frac{1}{2\gamma}} \right\rfloor$

$$\inf_{t\in[-k,k]} |\bar{\mathrm{w}}(t)||\,\mathrm{M}_U(t)| \geq \mathfrak{C}_2 \exp(-\mathfrak{c}_l k^{2\gamma})(\log k)^{-\frac{1}{2}-\rho}$$
$$\geq 6m^{-\frac{1}{2}}\mathfrak{C}_1(\log m)^{\frac{1}{2}} \geq 6\tau_m m^{-\frac{1}{2}}.$$

Therefore, exploiting eq. (41), i.e. $k_m \geq k$ and eq. (45), we obtain

$$\left\|\mathbb{1}_{[-k_m,k_m]^C}\,\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})} \leq \mathfrak{C}_5(c,a,s,L,\mathfrak{c}_l,\gamma,\eta_U,\rho)(\log m)^{-\frac{s-a}{\gamma}}.$$

Considering Theorem 4.6 and combining the last upper bound and the upper bounds in eqs. (25) and (29) in the proofs of Corollaries 2.5 and 3.6, respectively, we obtain the second claim, which completes the proof. □

Remark 4.2. Let us briefly revisit the choices of $a = 0$ and $a = -1$, which provide upper bounds for the global estimation risk for the density f and the survival function S^X, respectively:

1. Density estimation: Choosing $a = 0$ for all $f \in \mathbb{W}^s(L)$ leads to

$$\mathbb{E}\left[\left\|\hat{f}_{\hat{k}} - f\right\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})}\right] \leq \mathfrak{C}_1\left(n^{-\frac{2s}{2\gamma+2s+1}} + \left(\frac{(\log m)^{2(1+\rho)}}{m}\right)^{\left(\frac{s}{\gamma}\right)} \vee m^{-1}\right)$$

if M_U is ordinary smooth, and

$$\mathbb{E}\left[\left\|\hat{f}_{\hat{k}} - f\right\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-1})}\right] \leq \mathfrak{C}_2\left((\log n)^{-\frac{s}{\gamma}} + (\log m)^{-\frac{s}{\gamma}}\right)$$

if M_U is super smooth.

2. Survival function estimation: Choosing $a = 1$ for all $f \in \mathbb{W}^s(L)$ leads to

$$\mathbb{E}\left[\left\|\widehat{S}^X_{\hat{k}} - S\right\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-3})}\right] \leq \mathfrak{C}_1\left(n^{-\frac{2(s+1)}{2\gamma+2s+1}} + \left(\frac{(\log m)^{2(1+\rho)}}{m}\right)^{\left(\frac{s+1}{\gamma}\right)} \vee m^{-1}\right)$$

if M_U is ordinary smooth, and

$$\mathbb{E}\left[\left\|\widehat{S}^X_{\hat{k}} - S\right\|^2_{\mathbb{L}^2_+(\mathrm{x}^{2c-3})}\right] \leq \mathfrak{C}_2\left((\log n)^{-\frac{s+1}{\gamma}} + (\log m)^{-\frac{s+1}{\gamma}}\right)$$

if M_U is super smooth.

The oracle- and fully data-driven rates in Corollaries 3.6 and 4.7, respectively, coincide in case $\mathfrak{s.s.}$ for all $s > a$ and in case $\mathfrak{o.s.}$ for $s-a > \gamma$. In other words we do not pay an additional prize for the fully data-driven selection of the dimension parameter. In case $\mathfrak{s.s.}$ for $s - a \leq \gamma$ the rates differ, but the fully data-driven rate features only a deterioration by a factor $(\log m)^{2(1+\rho)(a-s)/\gamma}$. □

5. Numerical study

In this section we are going to illustrate the performance of the data-driven estimation procedure as presented in section 4. Eventually, we want to highlight the behaviour of $\hat{f}_{\hat{k}}$ under the influence of an increase of additional measurements for different types of unknown probability functions f of X. Particularly, we are interested in four different densities f of X, we aim to estimate, namely

i) Gamma – distribution, $\Gamma(q,p)$:

$$f_1(x) = \frac{q^p}{\Gamma(p)} x^{p-1} \exp(-qx) \mathbb{1}_{\mathbb{R}_{>0}}(x)$$

with $q = 1$ and $p = 3$.

ii) Weibull – distribution, $\mathrm{Weib}(l,k)$:

$$f_2(x) = s \cdot k \cdot (s \cdot x)^{k-1} \exp(-(s \cdot x)^k) \mathbb{1}_{\mathbb{R}_{>0}}(x)$$

with $s = 1$ and $k = 3$.

iii) Beta – distribution, $\mathrm{Beta}(a,b)$:

$$f_3(x) = \frac{1}{\mathrm{B}(a,b)} x^{a-1} (1-x)^{b-1} \mathbb{1}_{(0,1)}(x)$$

with $a = 10$ and $b = 5$.

iv) Log-Normal – distribution, $\mathrm{LogN}(\mu, \sigma^2)$:

$$f_4(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\Big(- \frac{(\ln(x) - \mu)^2}{2\sigma^2} \Big) \mathbb{1}_{\mathbb{R}_{>0}}(x)$$

with $\mu = 0$ and $\sigma^2 = 1$.

Moreover, as the unknown error distribution of U, we consider a Pareto distribution, $\mathrm{Pareto}(l, x_{\min})$,

$$f^U(x) = \frac{l x_{\min}^l}{x^{l+1}} \mathbb{1}_{[x_{\min},\infty)}(x)$$

with $l = 1$ and $x_{\min} = 1$. Elementary computations show that this choice of error distribution actually satisfies the ordinary smoothness conditions ($\mathfrak{o.s.}$) with $\gamma = 1$. In a first place we document how an increasing number of additional measurements m has an impact on the statistical behaviour of $\hat{f}_{\hat{k}}$. To do so, we consider first f_1 as target density and generate a sample $\{Y_i\}_{i \in [\![1000]\!]}$ following the law of $Y = X \cdot U$ with independent $X \sim \Gamma(1,3)$ and $U \sim \mathrm{Pareto}(1,1)$. Moreover, we have sampled $m \in \{100, 1000, 4000\}$ additional observations of $U \sim \mathrm{Pareto}(1,1)$. With those samples we have computed the data driven choice $\hat{k}$ and afterwards $\hat{f}_{\hat{k}}$ according to the estimation strategies presented in section 3 and section 4, where we have chosen $c = \frac{1}{2}$ and 0.3 as constant in the penalty. As note by several authors (see for instance [17]) the constant 48 in (35), though convenient for deriving the theory, is far too large in practise. In order to capture the randomness, we have repeated this procedure for $N = 500$ Monte-Carlo iterations, meaning we have computed a family of data driven spectral cut-off density estimators $\{\hat{f}^j_{\hat{k}_j}\}_{j \in [\![N]\!]}$. For a direct comparison to the situation of knowing the error density f^U, we have also computed a family $\{\tilde{f}^j_{\tilde{k}_j}\}_{j \in [\![N]\!]}$ of spectral cut-off density estimators of f_1 as presented in section 2 with $c = \frac{1}{2}$ and 0.6 as constant in the penalty. The true density f_1, the family of estimators,

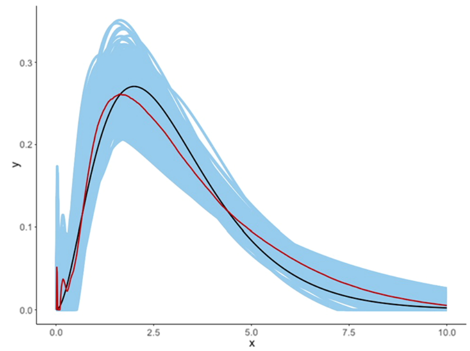

FIG 1. *Unknown error distribution, $n = 1000$, $m = 100$. Black line: true density f_1, Blue lines: N Monte-Carlo estimations of f_1, Red Line: point-wise median.* eMISE $= 0.00575$.

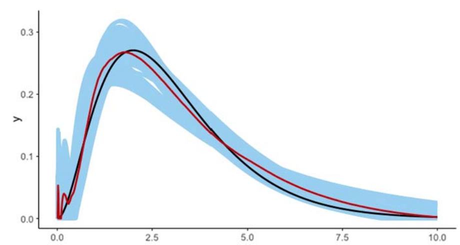

FIG 2. *Unknown error distribution, $n = 1000$, $m = 1000$. Black line: true density f_1, Blue lines: N Monte-Carlo estimations of f_1, Red Line: point-wise median.* eMISE $= 0.00458$.

as well as a point-wise computed median are depict in fig. 1 – fig. 4, where also the corresponding empirical mean integrated squared errors (eMISE) are stated. As the theory indicates, the estimation becomes more accurate for an increasing sample size $m = 100$ to 1000, while for $m = 1000$ to $m = 4000$ there is no significant improvement. And the accuracy corresponds nearly to the case of a known error density. Secondly, we illustrate the behaviour of $\widehat{f}_{\widehat{k}}$ for a fixed number of samples n and m, but for different target densities of X, namely for $f_1, \dots, f_4$. We fix $n = m = 2000$ and sample n observations of Y^i, $i \in [\![4]\!]$ according to the relation $Y^i = X^i \cdot U$, where X^i follows the law of f_i, $i \in [\![4]\!]$ and U is still Pareto$(1, 1)$ distributed. Again, given m additional measurements as i.i.d copies of U we compute afterwards $\widehat{k}$ as well as $\widehat{f}_{\widehat{k}}$ as before, following the definitions in section 3 and section 4 with choosing again $c = \frac{1}{2}$ and constant 0.3 for the penalty. Repeating this procedure for $N := 500$ Monte-Carlo iterations we obtain four families of estimators $\{\widehat{f}^{i,j}_{\widehat{k}_{i,j}}\}_{i \in [\![4]\!], j \in [\![N]\!]}$. The results as well as the empirical mean integrated squared error can be found in fig. 5 – fig. 8.

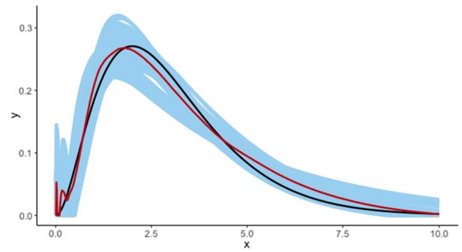

Fig 3. *Unknown error distribution, $n = 1000$, $m = 4000$. Black line: true density f_1, Blue lines: N Monte-Carlo estimations of f_1, Red Line: Point-wise median.* eMISE = 0.00449.

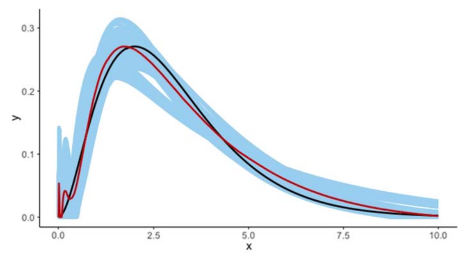

Fig 4. *Known error distribution, $n = 1000$. Black line: true density f_1, Blue lines: N Monte-Carlo estimations of f_1, Red Line: point-wise median.* eMISE = 0.00299.

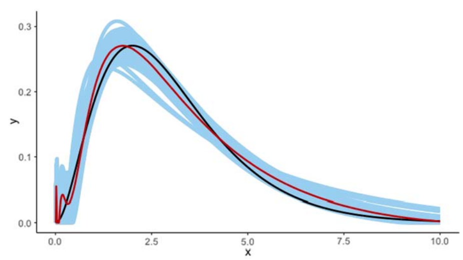

Fig 5. *Unknown error distribution, $n = m = 2000$. Black line: true density f_1, Blue lines: N Monte-Carlo estimations of f_1, Red Line: point-wise median.* eMISE = 0.00184.

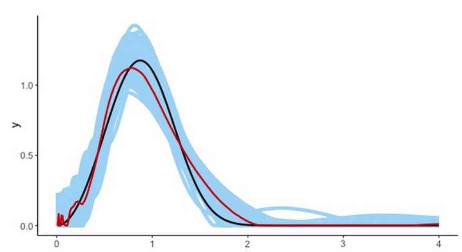

Fig 6. *Unknown error distribution, $n = m = 2000$. Black line: true density f_2, Blue lines: N Monte-Carlo estimations of f_2, Red Line: point-wise median.* eMISE = 0.0241.

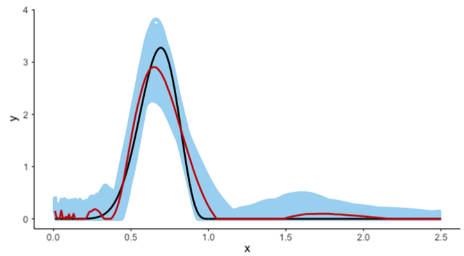

FIG 7. *Unknown error distribution,* $n = m = 2000$*. Black line: true density* f_3*, Blue lines: N Monte-Carlo estimations of* f_3*, Red Line: point-wise median.* eMISE $= 0.129$.

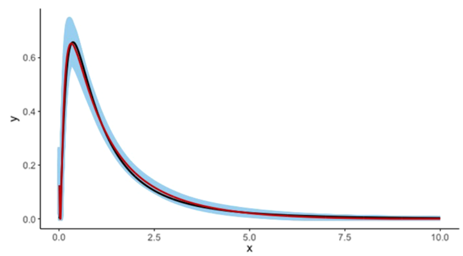

FIG 8. *Unknown error distribution,* $n = m = 2000$*. Black line: true density* f_4*, Blue lines: N Monte-Carlo estimations of* f_4*, Red Line: point-wise median.* eMISE $= 0.00179$.

Appendix A: Proofs of Section 2

In the end of this section we prove Proposition 2.3. The proof is based on Lemma A.2 below, which is formulated slightly more general, such that it can be reused again in the proof of Proposition 4.2 in Section 4. The next assertion and the subsequent Remark A.1 provide our key arguments in order to prove the concentration inequality in Lemma A.2. The next inequality is due to [28] and in this form for example given in [22].

Lemma A.1 (Talagrand's inequality). *Let* $(Z_i)_{i\in[\![n]\!]}$ *be independent* $\mathcal{Z}$*-valued random variables and let* $\{\nu_h : h \in \mathcal{H}\}$ *be countable class of Borel-measurable functions. For* $h \in \mathcal{H}$ *setting* $\overline{\nu}_h = n^{-1}\sum_{i\in[\![n]\!]}\{\nu_h(Z_i) - \mathbb{E}(\nu_h(Z_i))\}$ *we have*

$$\mathbb{E}\left[\left(\sup_{h\in\mathcal{H}}\{|\overline{\nu}_h|^2\} - 6\Psi^2\right)_+\right] \leq \mathfrak{C}_{\mathrm{tal}}\left[\frac{\tau}{n}\exp\left(\frac{-n\Psi^2}{6\tau}\right) + \frac{\psi^2}{n^2}\exp\left(\frac{-n\Psi}{100\psi}\right)\right] \tag{46}$$

for some universal numerical constants $\mathfrak{C}_{\mathrm{tal}} \in \mathbb{R}_{>0}$ *and where*

$$\sup_{h\in\mathcal{H}}\left\{\sup_{z\in\mathcal{Z}}\{|\nu_h(z)|\}\right\} \leq \psi, \qquad \mathbb{E}\left[\sup_{h\in\mathcal{H}}\{|\overline{\nu}_h|\}\right] \leq \Psi, \qquad \sup_{h\in\mathcal{H}}\left\{n\mathbb{E}\left[|\overline{\nu}_h|^2\right]\right\} \leq \tau.$$

Remark A.1. Consider an arbitrary density function $\mathrm{w} : \mathbb{R} \to \mathbb{R}_{\geq 0}$ such that $\mathbb{1}_{[-k,k]} \in \mathbb{L}^\infty(\mathrm{w})$ for each $k \in \mathbb{R}_{>0}$. For $k \in \mathbb{R}_{>0}$ let us briefly reconsider the orthogonal projection $(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\mathbb{1}_{[-k,k]}$. Introduce the subset $\mathbb{S}_k := \{h\mathbb{1}_{[-k,k]} : h \in \mathbb{L}^2(\mathrm{w})\}$ of $\mathbb{L}^2(\mathrm{w})$ and the unit ball $\mathbb{B}_k := \{h \in \mathbb{S}_k : \|h\|_{\mathbb{L}^2(\mathrm{w})} \leq 1\}$ in $\mathbb{S}_k$. Clearly, setting $\kappa : \mathbb{R}_{>0} \times \mathbb{R} \to \mathbb{C}$ with $(y,t) \mapsto \kappa(y,t) := y^{c-1}y^{\iota 2\pi t}$ we have $\widehat{\mathrm{M}}_Y(t) = \frac{1}{n}\sum_{i\in[\![n]\!]}\kappa(Y_i,t)$ and $\mathrm{M}_Y(t) = \mathbb{E}\big[\kappa(Y_i,t)\big]$ for each $t \in \mathbb{R}$. Introducing $\kappa_y : \mathbb{R} \to \mathbb{C}$ with $t \mapsto \kappa_y(t) := \kappa(y,t)$, $y \in \mathbb{R}_{>0}$, for each $h \in \mathbb{S}_k$ the function $\nu_h : \mathbb{R}_{>0} \to \mathbb{C}$ with $y \mapsto \nu_h(y) := \langle \kappa_y, h\rangle_{\mathbb{L}^2(\mathrm{w})}$ is Borel-measurable, where evidently $\frac{1}{n}\sum_{i\in[\![n]\!]}\nu_h(Y_i) = \langle \widehat{\mathrm{M}}_Y, h\rangle_{\mathbb{L}^2(\mathrm{w})}$ and hence $\overline{\nu}_h = \langle \widehat{\mathrm{M}}_Y - \mathrm{M}_Y, h\rangle_{\mathbb{L}^2(\mathrm{w})}$. Consequently, we obtain

$$\left\|(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\mathbb{1}_{[-k,k]}\right\|^2_{\mathbb{L}^2(\mathrm{w})} = \sup\{|\langle \widehat{\mathrm{M}}_Y - \mathrm{M}_Y, h\rangle_{\mathbb{L}^2(\mathrm{w})}|^2 : h \in \mathbb{B}_k\}$$
$$= \sup\{|\overline{\nu}_h|^2 : h \in \mathbb{B}_k\}.$$

Note that, the unit ball $\mathbb{B}_k$ is not a countable set, however, it contains a countable dense subset, say $\mathcal{B}_k$, since $\mathbb{L}^2(\mathrm{w})$ is separable. Exploiting the continuity of the inner product it is straightforward to see that

$$\left\|(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\mathbb{1}_{[-k,k]}\right\|^2_{\mathbb{L}^2(\mathrm{w})} = \sup\{|\overline{\nu}_h|^2 : h \in \mathbb{B}_k\} = \sup\{|\overline{\nu}_h|^2 : h \in \mathcal{B}_k\}.$$

The last identity provides the necessary argument to apply below Talagrand's inequality (A.1) where we need to calculate the three constants ψ, Ψ and τ. We note that the function $\kappa : \mathbb{R}_{>0} \times \mathbb{R} \to \mathbb{C}$ is not bounded. Therefore we decompose $\kappa = \kappa^{\mathrm{b}} + \kappa^{\mathrm{u}}$ into a bounded function $\kappa^{\mathrm{b}} : \mathbb{R}_{>0} \times \mathbb{R} \to \mathbb{C}$ with $(y,t) \mapsto \kappa^{\mathrm{b}}(y,t) := y^{c-1}\mathbb{1}_{(0,d)}(y^{c-1})y^{\iota 2\pi t}$ and an unbounded function $\kappa^{\mathrm{u}} : \mathbb{R}_{>0} \times \mathbb{R} \to \mathbb{C}$ with $(y,t) \mapsto \kappa^{\mathrm{u}}(y,t) := y^{c-1}\mathbb{1}_{[d,\infty)}(y^{c-1})y^{\iota 2\pi t}$ where $d := n^{1/3}$. For $\mathrm{j} \in \{\mathrm{b},\mathrm{u}\}$ setting $\widehat{\mathrm{M}}^{\mathrm{j}}_Y(t) = \frac{1}{n}\sum_{i\in[\![n]\!]}\kappa^{\mathrm{j}}(Y_i,t)$ and $\mathrm{M}^{\mathrm{j}}_Y(t) = \mathbb{E}\big[\kappa^{\mathrm{j}}(Y_i,t)\big]$ for each $t \in \mathbb{R}$ it follows $\widehat{\mathrm{M}}_Y - \mathrm{M}_Y = \widehat{\mathrm{M}}^{\mathrm{b}}_Y - \mathrm{M}^{\mathrm{b}}_Y + \widehat{\mathrm{M}}^{\mathrm{u}}_Y - \mathrm{M}^{\mathrm{u}}_Y$. Introducing further $\nu^{\mathrm{b}}_h : \mathbb{R}_{>0} \to \mathbb{C}$ with $y \mapsto \nu^{\mathrm{b}}_h(y) := \langle \kappa^{\mathrm{b}}_y, h\rangle_{\mathbb{L}^2(\mathrm{w})}$ and $\nu^{\mathrm{u}}_h : \mathbb{R}_{>0} \to \mathbb{C}$ with $y \mapsto \nu^{\mathrm{u}}_h(y) := \langle \kappa^{\mathrm{u}}_y, h\rangle_{\mathbb{L}^2(\mathrm{w})}$ we evidently have $\overline{\nu}_h = \overline{\nu}^{\mathrm{b}}_h + \overline{\nu}^{\mathrm{u}}_h$ and thus

$$\begin{aligned}\left\|(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\mathbb{1}_{[-k,k]}\right\|^2_{\mathbb{L}^2(\mathrm{w})} &= \sup\{|\overline{\nu}_h|^2 : h \in \mathcal{B}_k\} = \sup\{|\overline{\nu}^{\mathrm{b}}_h + \overline{\nu}^{\mathrm{u}}_h|^2 : h \in \mathcal{B}_k\}\\ &\leq 2\sup\{|\overline{\nu}^{\mathrm{b}}_h|^2 : h \in \mathcal{B}_k\} + 2\sup\{|\overline{\nu}^{\mathrm{u}}_h|^2 : h \in \mathcal{B}_k\}\\ &= 2\left\|(\widehat{\mathrm{M}}^{\mathrm{b}}_Y - \mathrm{M}^{\mathrm{b}}_Y)\mathbb{1}_{[-k,k]}\right\|^2_{\mathbb{L}^2(\mathrm{w})} + 2\left\|(\widehat{\mathrm{M}}^{\mathrm{u}}_Y - \mathrm{M}^{\mathrm{u}}_Y)\mathbb{1}_{[-k,k]}\right\|^2_{\mathbb{L}^2(\mathrm{w})}. \qquad (47)\end{aligned}$$

In Lemma A.2 below we bound the expectation of the first term on the right hand side with the help of Talagrand's inequality (Lemma A.1) and the expectation of the second term. □

Lemma A.2. *Let* $\mathrm{w} : \mathbb{R} \to \mathbb{R}_{\geq 0}$ *be a density function with* $\mathbb{1}_{[-k,k]} \in \mathbb{L}^\infty(\mathrm{w})$ *for each* $k \in \mathbb{R}_{>0}$. *Setting* $\sigma^2_Y := 1 + \mathbb{E}[Y_1^{2(c-1)}]$, $a_Y := \frac{6}{\sigma^2_Y}\left\|f^Y\right\|_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})}$, $k_Y := 1 \vee 3a^2_Y$,

$$\Delta^{\mathrm{w}}_k := \left\|\mathbb{1}_{[-k,k]}\right\|_{\mathbb{L}^\infty(\mathrm{w})} \text{ and } \delta^{\mathrm{w}}_k := \frac{\log(\Delta^{\mathrm{w}}_k \vee (k+2))}{\log(k+2)} \in \mathbb{R}_{\geq 1}, \quad \forall k \in \mathbb{R}_{\geq 1},$$

there exists an universal numerical constant $\mathfrak{C} \in \mathbb{R}_{>0}$ *such that for all* $n \in \mathbb{N}$ *and* $k \in \mathbb{R}_{\geq 1}$ *we have*

$$\mathbb{E}\left[\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y^{\mathrm{b}} - \mathrm{M}_Y^{\mathrm{b}})\right\|^2_{\mathbb{L}^2(\mathrm{w})} - 6\sigma_Y^2 \Delta_k^{\mathrm{w}} \delta_k^{\mathrm{w}} k n^{-1}\right)_+\right]$$
$$\leq n^{-1}\mathfrak{C}\left[(1 \vee \|f^Y\|^2_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})})(1 + \|\mathbb{1}_{[-k_Y,k_Y]}\|^2_{\mathbb{L}^2(\mathrm{w})})\frac{1}{a_Y}\exp\big(\frac{-k}{a_Y}\big) + n^{-4}k\Delta_k^{\mathrm{w}}\right] \tag{48}$$

and $\mathbb{E}\left[\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y^{\mathrm{u}} - \mathrm{M}_Y^{\mathrm{u}})\right\|^2_{\mathbb{L}^2(\mathrm{w})}\right] \leq 2\mathbb{E}\big[Y_1^{8(c-1)}\big]n^{-3}k\Delta_k^{\mathrm{w}}$.

Proof of Lemma A.2. Consider first the second claim. For each $t \in \mathbb{R}$ we have $\mathbb{E}\big[\widehat{\mathrm{M}}_Y^{\mathrm{u}}(t)\big] = \mathrm{M}_Y^{\mathrm{u}}(t) = \mathbb{E}\big[\kappa^{\mathrm{u}}(Y_1, t)\big]$ which in turn implies

$$\begin{aligned} n\mathbb{E}\big[|\widehat{\mathrm{M}}_Y^{\mathrm{u}}(t) - \mathrm{M}_Y^{\mathrm{u}}(t)|^2\big] &= \mathbb{V}\mathrm{ar}\big(\kappa^{\mathrm{u}}(Y_1,t)\big) \leq \mathbb{E}\big[|\kappa^{\mathrm{u}}(Y_1,t)|^2\big] \\ &= \mathbb{E}\big[Y_1^{2(c-1)}\mathbb{1}_{[d,\infty)}(Y_1^{c-1})\big] \leq d^{-2l}\mathbb{E}\big[Y_1^{2(c-1)(1+l)}\big] \end{aligned}$$

and thus making use of $d = n^{1/3}$ we obtain the claim, that is

$$\begin{aligned} \mathbb{E}\left[\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y^{\mathrm{u}} - \mathrm{M}_Y^{\mathrm{u}})\right\|^2_{\mathbb{L}^2(\mathrm{w})}\right] &\leq \tfrac{1}{nd^6}\mathbb{E}\big[Y_1^{2(c-1)(1+3)}\big]\,\|\mathbb{1}_{[-k,k]}\|^2_{\mathbb{L}^2(\mathrm{w})} \\ &\leq n^{-3}\mathbb{E}\big[Y_1^{8(c-1)}\big]2k\Delta_k^{\mathrm{w}}. \end{aligned}$$

Secondly consider (48). Given the identity (see Remark A.1)

$$\left\|(\widehat{\mathrm{M}}_Y^{\mathrm{b}} - \mathrm{M}_Y^{\mathrm{b}})\mathbb{1}_{[-k,k]}\right\|^2_{\mathbb{L}^2(\mathrm{w})} = \sup\{|\overline{\nu}_h^{\mathrm{b}}|^2 : h \in \mathcal{B}_k\}$$

we intent to apply Talagrand's inequality (Lemma A.1) where we need to calculate the three quantities ψ, Ψ and τ. Consider ψ first. Evidently, for each $k \in \mathbb{R}_{\geq 1}$ we have

$$\begin{aligned} \sup_{h\in\mathcal{B}_k}\left\{\sup_{y\in\mathbb{R}_{>0}}\big\{|\nu_h^{\mathrm{b}}(y)|^2\big\}\right\} &= \sup_{y\in\mathbb{R}_{>0}}\left\{\sup_{h\in\mathcal{B}_k}\left\{|\langle\kappa_y^{\mathrm{b}}, h\rangle_{\mathbb{L}^2(\mathrm{w})}|^2\right\}\right\} \\ &= \sup_{y\in\mathbb{R}_{>0}}\left\{\left\|\kappa_y^{\mathrm{b}}\mathbb{1}_{[-k,k]}\right\|^2_{\mathbb{L}^2(\mathrm{w})}\right\} \leq d^2\,\|\mathbb{1}_{[-k,k]}\|^2_{\mathbb{L}^2(\mathrm{w})} \leq 2d^2k\Delta_k^{\mathrm{w}} =: \psi^2. \end{aligned}$$

Consider next Ψ. Evidently, for each $t \in \mathbb{R}$ we have

$$n\mathbb{E}\big[|\widehat{\mathrm{M}}_Y^{\mathrm{b}}(t) - \mathrm{M}_Y^{\mathrm{b}}(t)|^2\big] = \mathbb{V}\mathrm{ar}\big(\kappa^{\mathrm{b}}(Y_1,t)\big) \leq \mathbb{E}\big[|\kappa^{\mathrm{b}}(Y_1,t)|^2\big] \leq \mathbb{E}\big[Y_1^{2(c-1)}\big] \leq \sigma_Y^2$$

which for each $k \in \mathbb{R}_{\geq 1}$ implies

$$\mathbb{E}\left[\sup_{h\in\mathcal{H}}|\overline{\nu}_h^{\mathrm{b}}|^2\right] = \mathbb{E}\left[\left\|(\widehat{\mathrm{M}}_Y^{\mathrm{b}} - \mathrm{M}_Y^{\mathrm{b}})\mathbb{1}_{[-k,k]}\right\|^2_{\mathbb{L}^2(\mathrm{w})}\right]$$

$$\le \tfrac{1}{n}\sigma_Y^2 \left\|\mathbb{1}_{[-k,k]}\right\|_{\mathbb{L}^2(\mathrm{w})}^2 \le \tfrac{1}{n}\sigma_Y^2 2k\Delta_k^{\mathrm{w}}\delta_k^{\mathrm{w}} =: \Psi^2.$$

Finally, consider τ. For each $h \in \mathcal{B}_k$ we have

$$n\mathbb{E}\big[|\overline{\nu}_h^{\mathrm{b}}|^2\big] = \mathbb{V}\mathrm{ar}\big[\langle \kappa^{\mathrm{b}}(Y_1,\cdot),h\rangle_{\mathbb{L}^2(\mathrm{w})}\big] \le \mathbb{E}\big[|\langle \kappa^{\mathrm{b}}(Y_1,\cdot),h\rangle_{\mathbb{L}^2(\mathrm{w})}|^2\big] \\ \le \mathbb{E}\big[|\langle \kappa(Y_1,\cdot),h\rangle_{\mathbb{L}^2(\mathrm{w})}|^2\big] = \mathbb{E}\big[|\nu_h|^2\big].$$

Since $h\mathrm{w} \in \mathbb{L}^2 \cap \mathbb{L}^1$ and $|\nu_h(y)|^2 = |\langle h,\kappa_y\rangle_{\mathbb{L}^2(\mathrm{w})}|^2 = |\,\mathrm{M}_{1-c}^{-1}[h\mathrm{w}](y)|^2$ for all $y \in \mathbb{R}_{>0}$ we have $\nu_h \in \mathbb{L}^2(\mathrm{x}^{1-2c})$. Consequently, using Parseval's identity we obtain

$$\|\nu_h\|_{\mathbb{L}^2(\mathrm{x}^{1-2c})}^2 = \|\mathrm{M}_{1-c}[\nu_h]\|_{\mathbb{L}^2}^2 = \|\mathrm{w}h\|_{\mathbb{L}^2}^2 \le \left\|\mathrm{w}\mathbb{1}_{[-k,k]}\right\|_{\mathbb{L}^\infty} \|h\|_{\mathbb{L}^2(\mathrm{w})}^2 \le \Delta_k^{\mathrm{w}}$$

which in turn for each $h \in \mathcal{B}_k$ implies

$$\mathbb{E}\big[|\nu_h|^2\big] \le \left\|f^Y\right\|_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})} \|\nu_h\|_{\mathbb{L}^2(\mathrm{x}^{1-2c})}^2 \le \left\|f^Y\right\|_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})} \Delta_k^{\mathrm{w}}.$$

On the other hand side for each $h \in \mathcal{B}_k$ we have also

$$\mathbb{E}\big[|\nu_h|^2\big] \le \left\|\mathbb{1}_{[-k,k]}\right\|_{\mathbb{L}^2(\mathrm{w})}^2 \mathbb{E}[Y_1^{2(c-1)}] \le \left\|\mathbb{1}_{[-k,k]}\right\|_{\mathbb{L}^2(\mathrm{w})}^2 \sigma_Y^2,$$

and hence

$$n\mathbb{E}\big[|\overline{\nu}_h^{\mathrm{b}}|^2\big] \le \left\|f^Y\right\|_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})} \Delta_k^{\mathrm{w}} \wedge \sigma_Y^2 \left\|\mathbb{1}_{[-k,k]}\right\|_{\mathbb{L}^2(\mathrm{w})}^2 =: \tau.$$

Given $a_Y = 6\left\|f^Y\right\|_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})}/\sigma_Y^2$ and $k_Y = 1 \vee 3a_Y^2$ for any $k \ge k_Y$ we have

$$\Delta_k^{\mathrm{w}} \exp\left(\frac{-\sigma_Y^2\delta_k^{\mathrm{w}}k}{6\left\|f^Y\right\|_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})}}\right) = \Delta_k^{\mathrm{w}} \exp\left(\tfrac{-\delta_k^{\mathrm{w}}}{a_Y}k\right) \le \exp(-\tfrac{\delta_k^{\mathrm{w}}}{a_Y}(k - a_Y\log(k+2))) \le 1$$

by exploiting that $x \ge a\log(x+2)$ for all $a > 0$ and $x \ge 1 \vee 3a^2$. Consequently, for any $k \ge k_Y$ we obtain (keep $\delta_k^{\mathrm{w}} \ge 1$ in mind)

$$\frac{\tau}{n}\exp\left(\frac{-n\Psi^2}{6\tau}\right) \le \frac{\left\|f^Y\right\|_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})}\Delta_k^{\mathrm{w}}}{n} \exp\left(\frac{-\sigma_Y^2\delta_k^{\mathrm{w}}k}{3\left\|f^Y\right\|_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})}}\right) \\ \le \frac{\left\|f^Y\right\|_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})}}{n} \exp\left(\frac{-k}{a_Y}\right)$$

while for any $k < k_Y$ we conclude (using again $\delta_k^{\mathrm{w}} \ge 1$)

$$\frac{\tau}{n}\exp\left(\frac{-n\Psi^2}{6\tau}\right) \le \frac{\sigma_Y^2\left\|\mathbb{1}_{[-k,k]}\right\|_{\mathbb{L}^2(\mathrm{w})}^2}{n} \exp\left(\frac{-\sigma_Y^2\delta_k^{\mathrm{w}}k}{3\left\|f^Y\right\|_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})}}\right) \\ \le \frac{\sigma_Y^2\left\|\mathbb{1}_{[-k_Y,k_Y]}\right\|_{\mathbb{L}^2(\mathrm{w})}^2}{n} \exp\left(\frac{-k}{a_Y}\right).$$

Combining both cases $k \geq k_Y$ and $k < k_Y$ we obtain for any $k \in \mathbb{R}_{\geq 1}$

$$\frac{\tau}{n} \exp\left(\frac{-n\Psi^2}{6\tau}\right) \leq \frac{\left\|f^Y\right\|_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})} + \sigma_Y^2 \left\|\mathbb{1}_{[-k_Y,k_Y]}\right\|^2_{\mathbb{L}^2(\mathrm{w})}}{n} \exp\big(\frac{-k}{a_Y}\big).$$

Evaluating the bound given by Talagrand's inequality (A.1) there exists an universal numerical constant $\mathfrak{C}_{\mathrm{tal}} \in \mathbb{R}_{>0}$ such that for each $k \in \mathbb{R}_{\geq 1}$ we have (keep $d = n^{1/3}$, $\sigma_Y^2 \geq 1$ and $\delta_k^{\mathrm{w}} \geq 1$ in mind)

$$\begin{aligned}
\mathbb{E}&\left[\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}^{\mathrm{b}}_Y - \mathrm{M}^{\mathrm{b}}_Y)\right\|^2_{\mathbb{L}^2(\mathrm{w})} - 6\sigma_Y^2 \Delta_k^{\mathrm{w}} \delta_k^{\mathrm{w}} k n^{-1}\right)_+\right] \\
&\leq n^{-1}\mathfrak{C}_{\mathrm{tal}}\Big[n^{-1/3} k \Delta_k^{\mathrm{w}} \exp\left(\frac{-n^{1/6}}{100}\right) \\
&\quad + (\left\|f^Y\right\|_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})} + \sigma_Y^2 \left\|\mathbb{1}_{[-k_Y,k_Y]}\right\|^2_{\mathbb{L}^2(\mathrm{w})}) \exp\big(\frac{-k}{a_Y}\big)\Big] \\
&\leq n^{-1}\mathfrak{C}\Big[n^{-4} k \Delta_k^{\mathrm{w}} n^{11/3} \exp\left(\frac{-n^{1/6}}{100}\right) \\
&\quad + (1 \vee \left\|f^Y\right\|^2_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})})(1 + \left\|\mathbb{1}_{[-k_Y,k_Y]}\right\|^2_{\mathbb{L}^2(\mathrm{w})}) \frac{1}{a_Y} \exp\big(\frac{-k}{a_Y}\big)\Big],
\end{aligned}$$

which together with $n^b \exp(-an^{1/c}) \leq (\frac{cb}{ae})^{cb}$ for all $a, b, c \in \mathbb{R}_{>0}$, and hence $n^{11/3} \exp(\frac{-n^{1/6}}{100}) \leq (1100)^{22}$ shows (48) and completes the proof. □

Proof of Proposition 2.3. From the decomposition (47) in Remark A.1 (with $\mathrm{w} = \mathrm{v}_U$) we obtain

$$\begin{aligned}
\mathbb{E}&\left[\max_{k \in [\![k_n]\!]}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\right\|^2_{\mathbb{L}^2(\mathrm{v}_U)} - 12\sigma_Y^2 \Delta_k^{\mathrm{v}_U} \delta_k^{\mathrm{v}_U} k n^{-1}\right)_+\right\}\right] \\
&\leq 2\mathbb{E}\left[\left\|\mathbb{1}_{[-k_n,k_n]}(\widehat{\mathrm{M}}^{\mathrm{u}}_Y - \mathrm{M}^{\mathrm{u}}_Y)\right\|^2_{\mathbb{L}^2(\mathrm{v}_U)}\right] \\
&\quad + 2\mathbb{E}\left[\max_{k \in [\![k_n]\!]}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}^{\mathrm{b}}_Y - \mathrm{M}^{\mathrm{b}}_Y)\right\|^2_{\mathbb{L}^2(\mathrm{v}_U)} - 6\sigma_Y^2 \Delta_k^{\mathrm{v}_U} \delta_k^{\mathrm{v}_U} k n^{-1}\right)_+\right\}\right],
\end{aligned}$$

where we bound the two right hand side terms separately with the help of Lemma A.2. Therewith we obtain

$$\begin{aligned}
\mathbb{E}&\left[\max_{k \in [\![k_n]\!]}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\right\|^2_{\mathbb{L}^2(\mathrm{v}_U)} - 12\sigma_Y^2 \Delta_k^{\mathrm{v}_U} \delta_k^{\mathrm{v}_U} k n^{-1}\right)_+\right\}\right] \\
&\leq n^{-1}\mathfrak{C}\Big[\mathbb{E}\big[Y_1^{8(c-1)}\big] n^{-2} k_n \Delta_{k_n}^{\mathrm{v}_U} + n^{-4} k_n^2 \Delta_{k_n}^{\mathrm{v}_U} \\
&\quad + (1 \vee \left\|f^Y\right\|^2_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})})(1 + \left\|\mathbb{1}_{[-k_Y,k_Y]}\right\|^2_{\mathbb{L}^2(\mathrm{v}_U)}) \sum_{k \in [\![k_n]\!]} \frac{1}{a_Y} \exp\big(\frac{-k}{a_Y}\big)\Big]. \qquad (49)
\end{aligned}$$

Exploiting that $\sum_{k\in\mathbb{N}}\exp(-k/a_Y)\leq a_Y$ and the definition of $k_n\in[\![n^2]\!]$ we have

$$\mathbb{E}\left[\max_{k\in[\![k_n]\!]}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y-\mathrm{M}_Y)\right\|^2_{\mathbb{L}^2(\mathrm{v}_U)}-12\sigma_Y^2\Delta_k^{\mathrm{v}_U}\delta_k^{\mathrm{v}_U}kn^{-1}\right)_+\right\}\right]$$
$$\leq n^{-1}\mathfrak{C}\left[(1\vee\mathbb{E}\big[Y_1^{8(c-1)}\big])\Delta_1^{\mathrm{v}_U}+(1\vee\left\|f^Y\right\|^2_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})})(1+\left\|\mathbb{1}_{[-k_Y,k_Y]}\right\|^2_{\mathbb{L}^2(\mathrm{v}_U)})\right],$$

which together with $\mathbb{E}\big[Y_1^{8(c-1)}\big]=\left\|f^Y\right\|_{\mathbb{L}^1(\mathrm{x}^{8(c-1)})}$, $\left\|\mathbb{1}_{[-k_Y,k_Y]}\right\|^2_{\mathbb{L}^2(\mathrm{v}_U)}\leq 2k_Y\Delta_{k_Y}^{\mathrm{v}_U}$, $\Delta_1^{\mathrm{v}_U}\leq k_Y\Delta_{k_Y}^{\mathrm{v}_U}$ and the definition of η_Y shows the claim and completes the proof. □

Appendix B: Proofs of Section 3

Lemma B.1. *There exists an universal numerical constant $\mathfrak{C}\in\mathbb{R}_{\geq 1}$ such that for any $k\in\mathbb{R}_{>0}$ we have*

$$\mathbb{E}\left[\left\|\mathbb{1}_{[-k,k]}\widehat{\mathrm{M}}_U^\dagger\mathbb{1}_{\mathfrak{M}}\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right]\leq 4\big(1\vee\mathbb{E}[U_1^{2(c-1)}]\big)\left\|\mathbb{1}_{[-k,k]}\,\mathrm{M}_U^\dagger\right\|^2_{\mathbb{L}^2(\mathrm{v})},$$
$$\mathbb{E}\left[\left\|\mathbb{1}_{[-k,k]}\,\mathrm{M}_X\,\mathbb{1}_{\mathfrak{M}^C}\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right]$$
$$\leq 4\big(1\vee\mathbb{E}[U_1^{2(c-1)}]\big)\left\|\mathbb{1}_{[-k,k]}\,\mathrm{M}_X\,\big(1\vee|\,\mathrm{M}_U\,|^2(m\wedge n)\big)^{-1/2}\right\|^2_{\mathbb{L}^2(\mathrm{v})},$$
$$\mathbb{E}\left[\left\|\mathbb{1}_{[-k,k]}\,\mathrm{M}_X(\mathrm{M}_U-\widehat{\mathrm{M}}_U)\widehat{\mathrm{M}}_U^\dagger\mathbb{1}_{\mathfrak{M}}\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right]$$
$$\leq 4\big(1\vee\mathfrak{C}\,\mathbb{E}[U_1^{4(c-1)}]\big)\left\|\mathbb{1}_{[-k,k]}\,\mathrm{M}_X\,\big(1\vee|\,\mathrm{M}_U\,|^2m\big)^{-1/2}\right\|^2_{\mathbb{L}^2(\mathrm{v})}.$$

Proof of Lemma B.1. We start our proof with the observation that for each $t\in\mathbb{R}$ we have $\mathbb{E}[\widehat{\mathrm{M}}_U(t)]=\mathrm{M}_U(t)$, $m\mathbb{E}\big[|\widehat{\mathrm{M}}_U(t)-\mathrm{M}_U(t)|^2\big]\leq\mathbb{E}[U_1^{2(c-1)}]$, and $m^2\mathbb{E}\big[|\widehat{\mathrm{M}}_U(t)-\mathrm{M}_U(t)|^4\big]\leq\mathfrak{C}\,\mathbb{E}[U_1^{4(c-1)}]$ for some universal numerical constant $\mathfrak{C}\in\mathbb{R}_{\geq 1}$ by applying Theorem 2.10 in [27]. We use those bounds without further reference. Below we show that for each $t\in\mathbb{R}$ we have

$$\mathbb{E}\left[|\,\mathrm{M}_U(t)\widehat{\mathrm{M}}_U^\dagger(t)|^2\mathbb{1}_{\mathfrak{M}}(t)\right]\leq 4\big(1\vee\mathbb{E}[U_1^{2(c-1)}]\big)\tag{50}$$
$$\mathbb{E}\left[\mathbb{1}_{\mathfrak{M}^C}(t)\right]=\mathbb{P}\left((n\wedge m)|\widehat{\mathrm{M}}_U(t)|^2<1\right)$$
$$\leq 4\big(1\vee\mathbb{E}[U_1^{2(c-1)}]\big)\big(1\vee|\,\mathrm{M}_U(t)|^2(m\wedge n)\big)^{-1}\tag{51}$$
$$\mathbb{E}\left[|\,\mathrm{M}_U(t)-\widehat{\mathrm{M}}_U(t)|^2|\widehat{\mathrm{M}}_U^\dagger(t)|^2\mathbb{1}_{\mathfrak{M}}(t)\right]$$
$$\leq 4\big(1\vee\mathfrak{C}\,\mathbb{E}[U_1^{4(c-1)}]\big)\big)(1\vee|\,\mathrm{M}_U(t)|^2m)^{-1}\tag{52}$$

Evidently, by applying Fubini's theorem from the last bounds follow immediatly Lemma B.1. It remains to show (50-52). Let $t \in \mathbb{R}$ be fixed. Consider (50) first. Exploiting $|\widehat{\mathrm{M}}_U(t)\widehat{\mathrm{M}}_U^{\dagger}(t)|^2 \mathbb{1}_{\mathfrak{M}}(t) = \mathbb{1}_{\mathfrak{M}}(t)$ and $2|\widehat{\mathrm{M}}_U(t)|^2 + 2|\mathrm{M}_U(t) - \widehat{\mathrm{M}}_U(t)|^2 \geq |\mathrm{M}_U(t)|^2$ we have

$$\begin{aligned}
\mathbb{E}\left[|\mathrm{M}_U(t)\widehat{\mathrm{M}}_U^{\dagger}(t)|^2 \mathbb{1}_{\mathfrak{M}}(t)\right] &\leq 2\mathbb{E}\left[|\mathrm{M}_U(t) - \widehat{\mathrm{M}}_U(t)|^2 |\widehat{\mathrm{M}}_U^{\dagger}(t)|^2 \mathbb{1}_{\mathfrak{M}}(t) + \mathbb{1}_{\mathfrak{M}}(t)\right] \\
&\leq 2\left((m \wedge n)\mathbb{E}[|\mathrm{M}_U(t) - \widehat{\mathrm{M}}_U(t)|^2] + 1\right) \\
&\leq 2\left(\mathbb{E}[U_1^{2(c-1)}] + 1\right) \leq 4\left(1 \vee \mathbb{E}[U_1^{2(c-1)}]\right)
\end{aligned}$$

which shows (50). Secondly, (51) is trivially satisfied if $1 \leq 4(1 \vee \mathbb{E}[U_1^{2(c-1)}])(1 \vee |\mathrm{M}_U(t)|^2 (n \wedge m))^{-1}$. Otherwise, from $(1 \vee |\mathrm{M}_U(t)|^2 (n \wedge m)) > 4(1 \vee \mathbb{E}[U_1^{2(c-1)}]) \geq 4$ follows $(n \wedge m)^{-1} < |\mathrm{M}_U(t)|^2/4$ which using Markov's inequality implies (51), that is

$$\begin{aligned}
\mathbb{P}\left(|\widehat{\mathrm{M}}_U(t)|^2 < (n \wedge m)^{-1}\right) &\leq \mathbb{P}\left(|\widehat{\mathrm{M}}_U(t) - \mathrm{M}_U(t)| > |\mathrm{M}_U(t)|/2\right) \\
&\leq 4\mathbb{E}\big[|\widehat{\mathrm{M}}_U(t) - \mathrm{M}_U(t)|^2\big] |\mathrm{M}_U(t)|^{-2} \\
&\leq 4\mathbb{E}[U_1^{2(c-1)}](|\mathrm{M}_U(t)|^2 m)^{-1} \\
&\leq 4(1 \vee \mathbb{E}[U_1^{2(c-1)}])(1 \vee |\mathrm{M}_U(t)|^2 (n \wedge m))^{-1}.
\end{aligned}$$

Finally, consider (52). Evidently, we have on the one hand

$$\mathbb{E}\left[|\mathrm{M}_U(t) - \widehat{\mathrm{M}}_U(t)|^2 |\widehat{\mathrm{M}}_U^{\dagger}(t)|^2 \mathbb{1}_{\mathfrak{M}}(t)\right] \leq (n \wedge m) m^{-1} \mathbb{E}[U_1^{2(c-1)}] \leq \mathbb{E}[U_1^{2(c-1)}]$$

while on the other hand

$$\begin{aligned}
&\mathbb{E}\left[|\mathrm{M}_U(t) - \widehat{\mathrm{M}}_U(t)|^2 |\widehat{\mathrm{M}}_U^{\dagger}(t)|^2 \mathbb{1}_{\mathfrak{M}}(t)\right] \\
&\leq 2\mathbb{E}\left[|\mathrm{M}_U(t) - \widehat{\mathrm{M}}_U(t)|^2 |\widehat{\mathrm{M}}_U^{\dagger}(t)|^2 \mathbb{1}_{\mathfrak{M}}(t) \Big(\frac{|\mathrm{M}_U(t) - \widehat{\mathrm{M}}_U(t)|^2}{|\mathrm{M}_U(t)|^2} + \frac{|\widehat{\mathrm{M}}_U(t)|^2}{|\mathrm{M}_U(t)|^2}\Big)\right] \\
&\qquad \leq 2\frac{(m \wedge n)\mathbb{E}\big[|\mathrm{M}_U(t) - \widehat{\mathrm{M}}_U(t)|^4\big]}{|\mathrm{M}_U(t)|^2} + 2\frac{\mathbb{E}\big[|\mathrm{M}_U(t) - \widehat{\mathrm{M}}_U(t)|^2\big]}{|\mathrm{M}_U(t)|^2} \\
&\qquad\qquad \leq 2(\mathfrak{C}\mathbb{E}[U_1^{4(c-1)}] + \mathbb{E}[U_1^{2(c-1)}])(|\mathrm{M}_U(t)|^2 m)^{-1}.
\end{aligned}$$

Combining both bounds we obtain (52), which completes the proof. □

Appendix C: Proofs of Section 4

We first recall an inequality due to [28] which in this form for example is stated by [7] in equation (5.13) in Corollary 2. We make use of it in the proof of Proposition 4.5 below.

Lemma C.1 (Talagrand's inequality)**.** *Let $(Z_i)_{i\in[\![m]\!]}$ be independent and identically distributed $\mathcal{Z}$-valued random variables and let $\{\nu_t : t \in \mathcal{T}\}$ be countable class of Borel-measurable functions. For $t \in \mathcal{T}$ setting*

$$\overline{\nu}_t = m^{-1} \sum_{i\in[\![m]\!]} \{\nu_t(Z_i) - \mathbb{E}(\nu_t(Z_i))\}$$

we have

$$\mathbb{P}\left[\sup_{t\in\mathcal{T}} |\overline{\nu}_t| \geq 2\Psi + \kappa\right] \leq 3\exp\left(-\mathfrak{C}_{\mathrm{tal}} m\big(\frac{\kappa^2}{\tau} \wedge \frac{\kappa}{\psi}\big)\right) \tag{53}$$

for some universal numerical constant $\mathfrak{C}_{\mathrm{tal}} \in \mathbb{R}_{>0}$ and where

$$\sup_{t\in\mathcal{T}}\left\{\sup_{z\in\mathcal{Z}}\{|\nu_t(z)|\}\right\} \leq \psi, \qquad \mathbb{E}\left[\sup_{t\in\mathcal{T}}|\overline{\nu}_t|\right] \leq \Psi, \qquad \sup_{t\in\mathcal{T}}\left\{\mathbb{E}\big[|\nu_t(Z_1)|^2\big]\right\} \leq \tau.$$

Proof of Lemma 4.1. Let $k_\circ \in \mathcal{K} := [\![k_n]\!]$ be arbitrary but fixed. Introduce $\widehat{\mathrm{M}}_X := \widehat{\mathrm{M}}_Y \widehat{\mathrm{M}}_U^{\dagger} \mathbb{1}_{\mathfrak{M}}$ and $\breve{\mathrm{M}}_X := \mathrm{M}_Y\, \widehat{\mathrm{M}}_U^{\dagger} \mathbb{1}_{\mathfrak{M}} = \mathrm{M}_X\, \mathrm{M}_U\, \widehat{\mathrm{M}}_U^{\dagger} \mathbb{1}_{\mathfrak{M}}$, for each $k \in \mathbb{R}_{>0}$ we write shortly $\widehat{\mathrm{M}}_X^k := \widehat{\mathrm{M}}_X \mathbb{1}_{[-k,k]}$, $\breve{\mathrm{M}}_X^k := \breve{\mathrm{M}}_X \mathbb{1}_{[-k,k]}$ and $\mathrm{M}_X^k := \mathrm{M}_X\, \mathbb{1}_{[-k,k]}$. Consider the disjoint decomposition $\mathcal{K} = \mathcal{K}_{<k_\circ} \cup \mathcal{K}_{\geq k_\circ}$ where $\mathcal{K}_{<k_\circ} := \{k \in \mathcal{K} : k < k_\circ\}$ and $\mathcal{K}_{\geq k_\circ} := \{k \in \mathcal{K} : k \geq k_\circ\}$, and similarly $\mathcal{K} = \mathcal{K}_{\leq k_\circ} \cup \mathcal{K}_{>k_\circ}$. Evidently, we have

$$\begin{aligned}\mathbb{1}_{[-\hat{k},\hat{k}]}\widehat{\mathrm{M}}_Y \widehat{\mathrm{M}}_U^{\dagger} \mathbb{1}_{\mathfrak{M}} - \mathrm{M}_X = \widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}} + (\breve{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{k_\circ})\mathbb{1}_{\hat{k}\in\mathcal{K}_{<k_\circ}} \\ + (\breve{\mathrm{M}}_X^{k_\circ} - \mathrm{M}_X)\mathbb{1}_{\hat{k}\in\mathcal{K}_{<k_\circ}} + (\breve{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X)\mathbb{1}_{\hat{k}\in\mathcal{K}_{\geq k_\circ}}\end{aligned}$$

which in turn implies

$$\begin{aligned}\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \leq 3\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 3\left\|\breve{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{k_\circ}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_{<k_\circ}} \\ + 3\big(\left\|\breve{\mathrm{M}}_X^{k_\circ} - \mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_{<k_\circ}} + \left\|\breve{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_{\geq k_\circ}}\big).\end{aligned}$$

Combining the last bound and the elementary estimate (keep in mind that $k_n := \max\mathcal{K}$)

$$\begin{aligned}\left\|\breve{\mathrm{M}}_X^{k_\circ} - \mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_{<k_\circ}} + \left\|\breve{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_{\geq k_\circ}} \\ \leq \left\|\breve{\mathrm{M}}_X^{k_n} - \mathrm{M}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + \left\|\mathrm{M}_X^{k_\circ} - \mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2\end{aligned}$$

we have

$$\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \leq 3\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 3\left\|\breve{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{k_\circ}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_{<k_\circ}} + 3\left\|\breve{\mathrm{M}}_X^{k_n} - \mathrm{M}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 3\left\|\mathrm{M}_X^{k_\circ} - \mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2$$

which together with the estimate

$$\begin{aligned}\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 &= \left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 (\mathbb{1}_{\hat{k}\in\mathcal{K}_{\leq k_\circ}} + \mathbb{1}_{\hat{k}\in\mathcal{K}_{>k_\circ}}) \\ &\leq \left\|\widehat{\mathrm{M}}_X^{k_\circ} - \breve{\mathrm{M}}_X^{k_\circ}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + \left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_{>k_\circ}}\end{aligned}$$

implies the upper bound

$$\begin{aligned}&\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X\right\|^2 \\ &\quad\leq 3\left\|\widehat{\mathrm{M}}_X^{k_\circ} - \breve{\mathrm{M}}_X^{k_\circ}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 3\left\|\mathrm{M}_X - \mathrm{M}_X^{k_\circ}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 3\left\|\breve{\mathrm{M}}_X^{k_n} - \mathrm{M}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \\ &\quad + 3\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_{>k_\circ}} + 3\left\|\breve{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{k_\circ}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_{<k_\circ}}.\end{aligned} \tag{54}$$

We bound separately the last two terms on the right hand side in (54) next. Consider $\left\|\breve{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{k_\circ}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_{<k_\circ}}$ first. Introduce the random decomposition $\mathcal{K}_{<k_\circ} = \mathcal{K}_- \cup \mathcal{K}_-^c$ with index set

$$\mathcal{K}_- := \{k \in \mathcal{K}_{<k_\circ} : \left\|\breve{\mathrm{M}}_X^{k_\circ} - \breve{\mathrm{M}}_X^{k}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 > 8\widehat{\sigma}_Y^2 \operatorname{pen}_{k_\circ}^{\hat{\mathrm{v}}}\} \tag{55}$$

and its complement $\mathcal{K}_-^c := \mathcal{K}_{<k_\circ} \setminus \mathcal{K}_-$. If $\mathcal{K}_-^c \neq \emptyset$ then for each $k \in \mathcal{K}_-^c$ we have

$$\left\|\breve{\mathrm{M}}_X^{k_\circ} - \breve{\mathrm{M}}_X^{k}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \leq \left\|\breve{\mathrm{M}}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \leq 2\left\|\breve{\mathrm{M}}_X^{k_n} - \mathrm{M}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 2\left\|\mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2$$

and also $\left\|\breve{\mathrm{M}}_X^{k_\circ} - \breve{\mathrm{M}}_X^{k}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \leq 8\widehat{\sigma}_Y^2 \operatorname{pen}_{k_\circ}^{\hat{\mathrm{v}}}$, which together imply

$$\begin{aligned}&\left\|\breve{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{k_\circ}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_-^c} \\ &\quad\leq \big(2\left\|\breve{\mathrm{M}}_X^{k_n} - \mathrm{M}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 2\left\|\mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\mho^c} + 8\widehat{\sigma}_Y^2 \operatorname{pen}_{k_\circ}^{\hat{\mathrm{v}}} \mathbb{1}_{\mho}\big)\mathbb{1}_{\hat{k}\in\mathcal{K}_-^c}.\end{aligned}$$

If $\mathcal{K}_- \neq \emptyset$ then for each $k \in \mathcal{K}_-$ we have

$$\tfrac{1}{2}\left\|\breve{\mathrm{M}}_X^{k_\circ} - \breve{\mathrm{M}}_X^{k}\right\|_{\mathbb{L}^2(\mathrm{v})}^2$$

$$\leq \left\| (\breve{\mathrm{M}}_X - \widehat{\mathrm{M}}_X)(\mathbb{1}_{[-k_\circ,k_\circ]} - \mathbb{1}_{[-k,k]}) \right\|^2_{\mathbb{L}^2(\mathrm{v})} + \left\| \widehat{\mathrm{M}}^{k_\circ}_X - \widehat{\mathrm{M}}^{k}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})}$$
$$\leq \left\| \breve{\mathrm{M}}^{k_\circ}_X - \widehat{\mathrm{M}}^{k_\circ}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} + \left\| \widehat{\mathrm{M}}^{k_\circ}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} - \left\| \widehat{\mathrm{M}}^{k}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})}$$

which using for the last estimate the definition (31) and (55) of $\hat{k}$ and $\mathcal{K}_-$, respectively, implies

$$\left\| \breve{\mathrm{M}}^{\hat{k}}_X - \breve{\mathrm{M}}^{k_\circ}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} \mathbb{1}_{\hat{k}\in\mathcal{K}_-}$$
$$\leq 4 \left\| \breve{\mathrm{M}}^{k_\circ}_X - \widehat{\mathrm{M}}^{k_\circ}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} \mathbb{1}_{\hat{k}\in\mathcal{K}_-} + 4\big(\left\| \widehat{\mathrm{M}}^{k_\circ}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} - \left\| \widehat{\mathrm{M}}^{\hat{k}}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} - 2\widehat{\sigma}^2_Y \,\mathrm{pen}^{\widehat{\mathrm{v}}}_{\hat{k}} \big)\mathbb{1}_{\hat{k}\in\mathcal{K}_-}$$
$$+ 4\big(2\widehat{\sigma}^2_Y \,\mathrm{pen}^{\widehat{\mathrm{v}}}_{\hat{k}} - \tfrac{1}{4} \left\| \breve{\mathrm{M}}^{k_\circ}_X - \breve{\mathrm{M}}^{\hat{k}}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} \big)\mathbb{1}_{\hat{k}\in\mathcal{K}_-}$$
$$\leq 4 \left\| \breve{\mathrm{M}}^{k_\circ}_X - \widehat{\mathrm{M}}^{k_\circ}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} \mathbb{1}_{\hat{k}\in\mathcal{K}_-}.$$

Combining both cases $\hat{k} \in \mathcal{K}_-$ and $\hat{k} \in \mathcal{K}^c_-$ we obtain the bound

$$\left\| \breve{\mathrm{M}}^{\hat{k}}_X - \breve{\mathrm{M}}^{k_\circ}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} \mathbb{1}_{\hat{k}\in\mathcal{K}_{<k_\circ}}$$
$$\leq \mathbb{1}_{\hat{k}\in\mathcal{K}_{<k_\circ}} \Bigg(4 \left\| \breve{\mathrm{M}}^{k_\circ}_X - \widehat{\mathrm{M}}^{k_\circ}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} + 2 \left\| \breve{\mathrm{M}}^{k_n}_X - \mathrm{M}^{k_n}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})}$$
$$+ 2 \left\| \mathrm{M}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} \mathbb{1}_{\mho^c} + 8\widehat{\sigma}^2_Y \,\mathrm{pen}^{\widehat{\mathrm{v}}}_{k_\circ} \mathbb{1}_{\mho} \Bigg). \tag{56}$$

Secondly, consider the term $\left\| \widehat{\mathrm{M}}^{\hat{k}}_X - \breve{\mathrm{M}}^{\hat{k}}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} \mathbb{1}_{\hat{k}\in\mathcal{K}_{>k_\circ}}$ in (54). Introduce the random decomposition $\mathcal{K}_{>k_\circ} = \mathcal{K}_+ \cup \mathcal{K}^c_+$ with index set

$$\mathcal{K}_+ := \{ k \in \mathcal{K}_{>k_\circ} : \left\| \widehat{\mathrm{M}}^{k}_X - \breve{\mathrm{M}}^{k}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} > 2\widehat{\sigma}^2_Y \,\mathrm{pen}^{\widehat{\mathrm{v}}}_{k_\circ} \} \tag{57}$$

and its complement $\mathcal{K}^c_+ := \mathcal{K}_{>k_\circ} \setminus \mathcal{K}_+$. If $\mathcal{K}^c_+ \neq \emptyset$ then for each $k \in \mathcal{K}^c_+$ we have $\left\| \widehat{\mathrm{M}}^{k}_X - \breve{\mathrm{M}}^{k}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} \leq \left\| \widehat{\mathrm{M}}^{k_n}_X - \breve{\mathrm{M}}^{k_n}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})}$ and $\left\| \widehat{\mathrm{M}}^{k}_X - \breve{\mathrm{M}}^{k}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} \leq 2\widehat{\sigma}^2_Y \,\mathrm{pen}^{\widehat{\mathrm{v}}}_{k_\circ}$, which together imply

$$\left\| \widehat{\mathrm{M}}^{\hat{k}}_X - \breve{\mathrm{M}}^{\hat{k}}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} \mathbb{1}_{\hat{k}\in\mathcal{K}^c_+} \leq \big(\left\| \widehat{\mathrm{M}}^{k_n}_X - \breve{\mathrm{M}}^{k_n}_X \right\|^2_{\mathbb{L}^2(\mathrm{v})} \mathbb{1}_{\mho^c} + 2\widehat{\sigma}^2_Y \,\mathrm{pen}^{\widehat{\mathrm{v}}}_{k_\circ} \mathbb{1}_{\mho} \big)\mathbb{1}_{\hat{k}\in\mathcal{K}^c_+}.$$

If $\mathcal{K}_+ \neq \emptyset$ then for each $k \in \mathcal{K}_+$ we have

$$- \left\| \widehat{\mathrm{M}}_X^{k_\circ} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 + \left\| \widehat{\mathrm{M}}_X^{k} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 - 2 \left\| \widehat{\mathrm{M}}_X^{k} - \breve{\mathrm{M}}_X^{k} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 - 2 \left\| \breve{\mathrm{M}}_X^{k} - \breve{\mathrm{M}}_X^{k_\circ} \right\|_{\mathbb{L}^2(\mathrm{v})}^2$$
$$= \left\| \widehat{\mathrm{M}}_X^{k} - \widehat{\mathrm{M}}_X^{k_\circ} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 - 2 \left\| \widehat{\mathrm{M}}_X^{k} - \breve{\mathrm{M}}_X^{k} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 - 2 \left\| \breve{\mathrm{M}}_X^{k} - \breve{\mathrm{M}}_X^{k_\circ} \right\|_{\mathbb{L}^2(\mathrm{v})}^2$$
$$\leq \left\| \widehat{\mathrm{M}}_X^{k} - \widehat{\mathrm{M}}_X^{k_\circ} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 - \left\| \widehat{\mathrm{M}}_X^{k} - \widehat{\mathrm{M}}_X^{k_\circ} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 = 0$$

which together with the definition (31) of $\hat{k}$ implies

$$\mathbb{1}_{\hat{k}\in\mathcal{K}_+} \big(- 2 \left\| \widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 - 2\widehat{\sigma}_Y^2 \,\mathrm{pen}_{k_\circ}^{\hat{\mathrm{v}}} + 2\widehat{\sigma}_Y^2 \,\mathrm{pen}_{\hat{k}}^{\hat{\mathrm{v}}} - 2 \left\| \breve{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{k_\circ} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \big)$$
$$\leq \mathbb{1}_{\hat{k}\in\mathcal{K}_+} \big(\left\| \widehat{\mathrm{M}}_X^{k_\circ} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 - 2\widehat{\sigma}_Y^2 \,\mathrm{pen}_{k_\circ}^{\hat{\mathrm{v}}} + 2\widehat{\sigma}_Y^2 \,\mathrm{pen}_{\hat{k}}^{\hat{\mathrm{v}}} - \left\| \widehat{\mathrm{M}}_X^{\hat{k}} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \big) \leq 0.$$

From the last elementary bound we obtain (keep the definition (57) of $\mathcal{K}_+$ and $\mathfrak{A} := \{\sigma_Y^2 \leq 2\widehat{\sigma}_Y^2\}$ in mind)

$$\begin{aligned}
&\left\| \widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_+} \mathbb{1}_{\mathfrak{A}} \\
&= \mathbb{1}_{\hat{k}\in\mathcal{K}_+} \mathbb{1}_{\mathfrak{A}} \big(4 \left\| \widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 - 2\widehat{\sigma}_Y^2 \,\mathrm{pen}_{\hat{k}}^{\hat{\mathrm{v}}} + 2 \left\| \breve{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{k_\circ} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \big) \\
&+ \mathbb{1}_{\hat{k}\in\mathcal{K}_+} \mathbb{1}_{\mathfrak{A}} \big(- \left\| \widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 2\widehat{\sigma}_Y^2 \,\mathrm{pen}_{k_\circ}^{\hat{\mathrm{v}}} \big) \\
&+ \mathbb{1}_{\hat{k}\in\mathcal{K}_+} \mathbb{1}_{\mathfrak{A}} \big(- 2 \left\| \widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 2\widehat{\sigma}_Y^2 \,\mathrm{pen}_{\hat{k}}^{\hat{\mathrm{v}}} \\
&\qquad\qquad - 2\widehat{\sigma}_Y^2 \,\mathrm{pen}_{k_\circ}^{\hat{\mathrm{v}}} - 2 \left\| \breve{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{k_\circ} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \big) \\
&\leq \mathbb{1}_{\hat{k}\in\mathcal{K}_+} \mathbb{1}_{\mathfrak{A}} \big(4 \left\| \widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 - 2\widehat{\sigma}_Y^2 \,\mathrm{pen}_{\hat{k}}^{\hat{\mathrm{v}}} + 2 \left\| \breve{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{k_\circ} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \big) \\
&\leq \mathbb{1}_{\hat{k}\in\mathcal{K}_+} \mathbb{1}_{\mathfrak{A}} \big(4 \left\| \widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 - \sigma_Y^2 \,\mathrm{pen}_{\hat{k}}^{\hat{\mathrm{v}}} + 2 \left\| \breve{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{k_\circ} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \big) \\
&\leq \mathbb{1}_{\hat{k}\in\mathcal{K}_+} \Big(4 \max_{k\in\mathcal{K}_+} \big\{ \big(\left\| \widehat{\mathrm{M}}_X^{k} - \breve{\mathrm{M}}_X^{k} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 - \tfrac{\sigma_Y^2}{4} \,\mathrm{pen}_{k}^{\hat{\mathrm{v}}} \big)_+ \big\} + 2 \left\| \breve{\mathrm{M}}_X^{k_n} - \breve{\mathrm{M}}_X^{k_\circ} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \Big).
\end{aligned}$$

which together with the elementary bounds

$$\left\| \breve{\mathrm{M}}_X^{k_n} - \breve{\mathrm{M}}_X^{k_\circ} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \leq 2 \left\| \breve{\mathrm{M}}_X^{k_n} - \mathrm{M}_X^{k_n} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 2 \left\| \mathrm{M}_X^{k_\circ} - \mathrm{M}_X \right\|_{\mathbb{L}^2(\mathrm{v})}^2$$

and $\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_+}\mathbb{1}_{\mathfrak{A}^c} \leq \left\|\widehat{\mathrm{M}}_X^{k_n} - \breve{\mathrm{M}}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_+}\mathbb{1}_{\mathfrak{A}^c}$ implies

$$\begin{aligned}\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_+} \leq \mathbb{1}_{\hat{k}\in\mathcal{K}_+}\Bigg(&\left\|\widehat{\mathrm{M}}_X^{k_n} - \breve{\mathrm{M}}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\mathfrak{A}^c} \\ &+ 4\max_{k\in\mathcal{K}_+}\left\{\Big(\left\|\widehat{\mathrm{M}}_X^{k} - \breve{\mathrm{M}}_X^{k}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 - \tfrac{\sigma_Y^2}{4}\operatorname{pen}_k^{\hat{\mathrm{v}}}\Big)_+\right\} \\ &+ 4\left\|\breve{\mathrm{M}}_X^{k_n} - \mathrm{M}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 4\left\|\mathrm{M}_X^{k_\circ} - \mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2\Bigg).\end{aligned}$$

Combining both cases $\hat{k}\in\mathcal{K}_+$ and $\hat{k}\in\mathcal{K}_+^c$ we obtain the bound

$$\begin{aligned}&\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_{>k_\circ}} \\ &\leq \mathbb{1}_{\hat{k}\in\mathcal{K}_+^c}\Big(\left\|\widehat{\mathrm{M}}_X^{k_n} - \breve{\mathrm{M}}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\mho^c} + 2\widehat{\sigma}_Y^2\operatorname{pen}_{k_\circ}^{\hat{\mathrm{v}}}\mathbb{1}_{\mho}\Big) \\ &+ \mathbb{1}_{\hat{k}\in\mathcal{K}_+}\Bigg(\left\|\widehat{\mathrm{M}}_X^{k_n} - \breve{\mathrm{M}}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\mathfrak{A}^c} + 4\max_{k\in\mathcal{K}_+}\Big\{\Big(\left\|\widehat{\mathrm{M}}_X^{k} - \breve{\mathrm{M}}_X^{k}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 - \tfrac{\sigma_Y^2}{4}\operatorname{pen}_k^{\hat{\mathrm{v}}}\Big)_+\Big\} \\ &\qquad + 4\left\|\breve{\mathrm{M}}_X^{k_n} - \mathrm{M}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 4\left\|\mathrm{M}_X^{k_\circ} - \mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2\Bigg) \\ &\leq \mathbb{1}_{\hat{k}\in\mathcal{K}_{>k_\circ}}\Bigg(\left\|\widehat{\mathrm{M}}_X^{k_n} - \breve{\mathrm{M}}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 (\mathbb{1}_{\mho^c} + \mathbb{1}_{\mathfrak{A}^c}) + 2\widehat{\sigma}_Y^2\operatorname{pen}_{k_\circ}^{\hat{\mathrm{v}}}\mathbb{1}_{\mho} \\ &\qquad + 4\max_{k\in\mathcal{K}_+}\Big\{\Big(\left\|\widehat{\mathrm{M}}_X^{k} - \breve{\mathrm{M}}_X^{k}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 - \tfrac{\sigma_Y^2}{4}\operatorname{pen}_k^{\hat{\mathrm{v}}}\Big)_+\Big\} \\ &\qquad + 4\left\|\breve{\mathrm{M}}_X^{k_n} - \mathrm{M}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 4\left\|\mathrm{M}_X^{k_\circ} - \mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2\Bigg). \end{aligned} \tag{58}$$

Making use of (56) and (58) we obtain

$$\begin{aligned}&\left\|\breve{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{k_\circ}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_-} + \left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \breve{\mathrm{M}}_X^{\hat{k}}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\hat{k}\in\mathcal{K}_{>k_\circ}} \\ &\quad\leq \mathbb{1}_{\hat{k}\in\mathcal{K}_{<k_\circ}}\Bigg(4\left\|\breve{\mathrm{M}}_X^{k_\circ} - \widehat{\mathrm{M}}_X^{k_\circ}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 \\ &\qquad + 2\left\|\breve{\mathrm{M}}_X^{k_n} - \mathrm{M}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 2\left\|\mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2\mathbb{1}_{\mho^c} + 8\widehat{\sigma}_Y^2\operatorname{pen}_{k_\circ}^{\hat{\mathrm{v}}}\mathbb{1}_{\mho}\Bigg) \\ &\quad + \mathbb{1}_{\hat{k}\in\mathcal{K}_{>k_\circ}}\Bigg(\left\|\widehat{\mathrm{M}}_X^{k_n} - \breve{\mathrm{M}}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 (\mathbb{1}_{\mho^c} + \mathbb{1}_{\mathfrak{A}^c}) + 2\widehat{\sigma}_Y^2\operatorname{pen}_{k_\circ}^{\hat{\mathrm{v}}}\mathbb{1}_{\mho} \\ &\qquad + 4\max_{k\in\mathcal{K}_+}\Big\{\Big(\left\|\widehat{\mathrm{M}}_X^{k} - \breve{\mathrm{M}}_X^{k}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 - \tfrac{\sigma_Y^2}{4}\operatorname{pen}_k^{\hat{\mathrm{v}}}\Big)_+\Big\} \\ &\qquad + 4\left\|\breve{\mathrm{M}}_X^{k_n} - \mathrm{M}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2 + 4\left\|\mathrm{M}_X^{k_\circ} - \mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2\Bigg)\end{aligned}$$

$$
\begin{aligned}
&\leq 4\left\|\mathrm{M}_X^{k_\circ}-\mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2+4\left\|\breve{\mathrm{M}}_X^{k_\circ}-\widehat{\mathrm{M}}_X^{k_\circ}\right\|_{\mathbb{L}^2(\mathrm{v})}^2+8\widehat{\sigma}_Y^2\operatorname{pen}_{k_\circ}^{\widehat{\mathrm{v}}}\mathbb{1}_{\mho}\\
&\quad+4\left\|\breve{\mathrm{M}}_X^{k_n}-\mathrm{M}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2+4\max_{k\in\mathcal{K}_+}\left\{\left(\left\|\widehat{\mathrm{M}}_X^{k}-\breve{\mathrm{M}}_X^{k}\right\|_{\mathbb{L}^2(\mathrm{v})}^2-\tfrac{\sigma_Y^2}{4}\operatorname{pen}_k^{\widehat{\mathrm{v}}}\right)_+\right\}\\
&\quad+\left\|\widehat{\mathrm{M}}_X^{k_n}-\breve{\mathrm{M}}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2(\mathbb{1}_{\mho^c}+\mathbb{1}_{\mathfrak{A}^c})+2\left\|\mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2\mathbb{1}_{\mho^c}
\end{aligned}
$$

which together with (54) implies the claim

$$
\begin{aligned}
&\left\|\widehat{\mathrm{M}}_X^{\hat{k}}-\mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2\\
&\quad\leq 3\left\|\widehat{\mathrm{M}}_X^{k_\circ}-\breve{\mathrm{M}}_X^{k_\circ}\right\|_{\mathbb{L}^2(\mathrm{v})}^2+3\left\|\mathrm{M}_X-\mathrm{M}_X^{k_\circ}\right\|_{\mathbb{L}^2(\mathrm{v})}^2+3\left\|\breve{\mathrm{M}}_X^{k_n}-\mathrm{M}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2\\
&\qquad+3\left\|\widehat{\mathrm{M}}_X^{\hat{k}}-\breve{\mathrm{M}}_X^{\hat{k}}\right\|_{\mathbb{L}^2(\mathrm{v})}^2\mathbb{1}_{\hat{k}\in\mathcal{K}_{>k_\circ}}+3\left\|\breve{\mathrm{M}}_X^{\hat{k}}-\breve{\mathrm{M}}_X^{k_\circ}\right\|_{\mathbb{L}^2(\mathrm{v})}^2\mathbb{1}_{\hat{k}\in\mathcal{K}_{<k_\circ}}\\
&\quad\leq 15\left\|\widehat{\mathrm{M}}_X^{k_\circ}-\breve{\mathrm{M}}_X^{k_\circ}\right\|_{\mathbb{L}^2(\mathrm{v})}^2+15\left\|\mathrm{M}_X-\mathrm{M}_X^{k_\circ}\right\|_{\mathbb{L}^2(\mathrm{v})}^2+24\widehat{\sigma}_Y^2\operatorname{pen}_{k_\circ}^{\widehat{\mathrm{v}}}\mathbb{1}_{\mho}\\
&\qquad+15\left\|\breve{\mathrm{M}}_X^{k_n}-\mathrm{M}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2+12\max_{k\in\mathcal{K}_+}\left\{\left(\left\|\widehat{\mathrm{M}}_X^{k}-\breve{\mathrm{M}}_X^{k}\right\|_{\mathbb{L}^2(\mathrm{v})}^2-\tfrac{\sigma_Y^2}{4}\operatorname{pen}_k^{\widehat{\mathrm{v}}}\right)_+\right\}\\
&\quad+3\left\|\widehat{\mathrm{M}}_X^{k_n}-\breve{\mathrm{M}}_X^{k_n}\right\|_{\mathbb{L}^2(\mathrm{v})}^2(\mathbb{1}_{\mho^c}+\mathbb{1}_{\mathfrak{A}^c})+6\left\|\mathrm{M}_X\right\|_{\mathbb{L}^2(\mathrm{v})}^2\mathbb{1}_{\mho^c}
\end{aligned}
$$

and completes the proof. □

Proof of Proposition 4.2. Since $\widehat{\mathrm{v}}:=|\widehat{\mathrm{M}}_U^{\dagger}\mathbb{1}_{\mathfrak{M}}|^2\mathrm{v}$ depends on the sample $\{U_j\}_{j\in[\![m]\!]}$ only, we apply the law of total expectation leading to

$$
\begin{aligned}
&\mathbb{E}\left[\max_{k\in[\![k_n]\!]}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y-\mathrm{M}_Y)\right\|_{\mathbb{L}^2(\widehat{\mathrm{v}})}^2-\sigma_Y^2\frac{\operatorname{pen}_k^{\widehat{\mathrm{v}}}}{4}\right)_+\right\}\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[\max_{k\in[\![k_n]\!]}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y-\mathrm{M}_Y)\right\|_{\mathbb{L}^2(\widehat{\mathrm{v}})}^2-\sigma_Y^2\frac{\operatorname{pen}_k^{\widehat{\mathrm{v}}}}{4}\right)_+\right\}\middle|\{U_j\}_{j\in[\![m]\!]}\right]\right].
\end{aligned}
$$

Evidently, conditioning on $\{U_j\}_{j\in[\![m]\!]}$ the density function $\widehat{\mathrm{v}}$ is deterministic, thus similar to the proof of (49) making again use of the decomposition (47) in Remark A.1 (with $\mathrm{w}=\widehat{\mathrm{v}}$) and applying Lemma A.2 we obtain

$$
\begin{aligned}
&\mathbb{E}\left[\max_{k\in[\![k_n]\!]}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y-\mathrm{M}_Y)\right\|_{\mathbb{L}^2(\widehat{\mathrm{v}})}^2-12\sigma_Y^2\Delta_k^{\widehat{\mathrm{v}}}\delta_k^{\mathrm{h}}kn^{-1}\right)_+\right\}\middle|\{U_j\}_{j\in[\![m]\!]}\right]\\
&\quad\leq n^{-1}\mathfrak{C}\Big[\mathbb{E}\big[Y_1^{8(c-1)}\big]n^{-2}k_n\Delta_{k_n}^{\widehat{\mathrm{v}}}+n^{-4}k_n^2\Delta_{k_n}^{\widehat{\mathrm{v}}}\\
&\qquad+(1\vee\left\|f^Y\right\|_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})}^2)(1+\left\|\mathbb{1}_{[-k_Y,k_Y]}\right\|_{\mathbb{L}^2(\widehat{\mathrm{v}})}^2)\sum_{k\in[\![k_n]\!]}\frac{1}{a_Y}\exp\left(\frac{-k}{a_Y}\right)\Big].
\end{aligned}
$$

Observing that $\Delta_k^{\widehat{\mathrm{v}}} \leq (m \wedge n)\Delta_k^{\mathrm{v}}$ and exploiting $\sum_{k\in\mathbb{N}} \exp(-k/a_Y) \leq a_Y$ as well as the definition (34) of $k_n \in [\![n]\!]$, we have

$$\mathbb{E}\left[\max_{k\in[\![k_n]\!]}\left\{\left(\left\|\mathbb{1}_{[-k,k]}(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\right\|^2_{\mathbb{L}^2(\widehat{\mathrm{v}})} - 12\sigma_Y^2\Delta_k^{\widehat{\mathrm{v}}}\delta_k^{\widehat{\mathrm{v}}}kn^{-1}\right)_+\right\}\Bigg|\{U_j\}_{j\in[\![m]\!]}\right]$$
$$\leq n^{-1}\mathfrak{C}\left[(1 \vee \mathbb{E}\big[Y_1^{8(c-1)}\big])\Delta_1^{\mathrm{v}} + (1 \vee \|f^Y\|^2_{\mathbb{L}^\infty(\mathrm{x}^{2c-1})})(1 + \left\|\mathbb{1}_{[-k_Y,k_Y]}\right\|^2_{\mathbb{L}^2(\widehat{\mathrm{v}})}))\right].$$

The last upper bound does not depend on the additional measurements $\{U_j\}_{j\in[\![m]\!]}$ up to the last norm, namely $\left\|\mathbb{1}_{[-k_Y,k_Y]}\right\|^2_{\mathbb{L}^2(\widehat{\mathrm{v}})} = \left\|\mathbb{1}_{[-k_Y,k_Y]}\widehat{\mathrm{M}}_U^\dagger\mathbb{1}_{\mathfrak{M}}\right\|^2_{\mathbb{L}^2(\mathrm{v})}$. Its expectation under the distribution of $\{U_j\}_{j\in[\![m]\!]}$ is bounded in Lemma B.1 as follows

$$\mathbb{E}\left[\left\|\mathbb{1}_{[-k_Y,k_Y]}\right\|^2_{\mathbb{L}^2(\widehat{\mathrm{v}})}\right] \leq 4\left(1 \vee \mathbb{E}[U_1^{2(c-1)}]\right)\left\|\mathbb{1}_{[-k_Y,k_Y]}\right\|^2_{\mathbb{L}^2(\mathrm{v})}.$$

Computing the total expectation together with $\left\|\mathbb{1}_{[-k_Y,k_Y]}\right\|^2_{\mathbb{L}^2(\mathrm{v})} \leq 2k_Y\Delta_{k_Y}^{\mathrm{v}}$, $\Delta_1^{\mathrm{v}} \leq k_Y\Delta_{k_Y}^{\mathrm{v}}$ and the definition of η_Y and η_U leads to the claim. □

Proof of Lemma 4.3. We start the proof with taking the expectation on both sides of the upper bound given in Lemma 4.1 and making use of the concentration inequality in Proposition 4.2, which for each $k_\circ \in [\![k_n]\!]$ leads to (similar to the proof of Lemma 3.2)

$$\begin{aligned}
\mathbb{E}&\left[\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right] \\
&\leq 15\frac{1}{n}\mathbb{E}\left[\left\|\mathbb{1}_{[-k_\circ,k_\circ]}\mathbb{V}_Y\widehat{\mathrm{M}}_U^\dagger\mathbb{1}_{\mathfrak{M}}\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right] + 15\left\|\mathbb{1}_{[-k_\circ,k_\circ]^C}\,\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})} \\
&\quad + 24\mathbb{E}\left[\widehat{\sigma}_Y^2\,\mathrm{pen}_{k_\circ}^{\widehat{\mathrm{v}}}\,\mathbb{1}_{\mho_{k_\circ}}\right] + 15\mathbb{E}\left[\left\|\mathbb{1}_{\mathfrak{M}^C}\,\mathrm{M}_X^{k_n}\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right] \\
&\quad + 15\mathbb{E}\left[\left\|\widehat{\mathrm{M}}_U^\dagger\mathbb{1}_{\mathfrak{M}}(\mathrm{M}_U - \widehat{\mathrm{M}}_U)\,\mathrm{M}_X^{k_n}\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right] + \mathfrak{C}\cdot\eta_Y(1 \vee \eta_U^2 k_Y\Delta_{k_Y}^{\mathrm{v}})\cdot n^{-1} \\
&\quad + 3\mathbb{E}\left[\left\|\mathbb{1}_{[-k_n,k_n]}(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\widehat{\mathrm{M}}_U^\dagger\mathbb{1}_{\mathfrak{M}}\right\|^2_{\mathbb{L}^2(\mathrm{v})}(\mathbb{1}_{\mho_{k_\circ}^C} + \mathbb{1}_{\mathfrak{A}^C})\right] \\
&\quad + 6\left\|\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})}\mathbb{P}(\mho_{k_\circ}^C).
\end{aligned}$$

The first, fourth and fifth term in the last upper bound we estimate with the help of Lemma B.1 (line by line as in Subsection 3.3 and using $\left\|\mathrm{M}_U^\dagger\,\mathbb{1}_{[-k_\circ,k_\circ]}\right\|^2_{\mathbb{L}^2(\mathrm{v})} \leq \Delta_{k_\circ}^{\mathrm{v}_U}\delta_{k_\circ}^{\mathrm{v}_U}k_\circ$ as well as the definition of η_U), which implies

$$\mathbb{E}\left[\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right]$$

$$\leq \mathfrak{C}\eta_U^4 \cdot \mathbb{E}[X_1^{2(c-1)}] \cdot \Delta_{k_\circ}^{\mathrm{v}_U} \delta_{k_\circ}^{\mathrm{v}_U} k_\circ n^{-1} + 15 \left\| \mathbb{1}_{[-k_\circ,k_\circ]^C} \mathrm{M}_X \right\|_{\mathbb{L}^2(\mathrm{v})}^2$$
$$+ 24\mathbb{E}\left[\widehat{\sigma}_Y^2 \operatorname{pen}_{k_\circ}^{\widehat{\mathrm{v}}} \mathbb{1}_{\mho_{k_\circ}}\right] + \mathfrak{C}\eta_U^4 \left\| \mathrm{M}_X \left(1 \vee |\,\mathrm{M}_U\,|^2 m\right)^{-1/2} \right\|_{\mathbb{L}^2(\mathrm{v})}^2$$
$$+ \mathfrak{C} \cdot \eta_Y (1 \vee \eta_U^2 k_Y \Delta_{k_Y}^{\mathrm{v}}) \cdot n^{-1}$$
$$+ 3\mathbb{E}\left[\left\| \mathbb{1}_{[-k_n,k_n]} (\widehat{\mathrm{M}}_Y - \mathrm{M}_Y) \widehat{\mathrm{M}}_U^{\dagger} \mathbb{1}_{\mathfrak{M}} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 (\mathbb{1}_{\mho_{k_\circ}^C} + \mathbb{1}_{\mathfrak{A}^C})\right]$$
$$+ 6 \left\| \mathrm{M}_X \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{P}(\mho_{k_\circ}^C).$$

Since $\Delta_{k_\circ}^{\widehat{\mathrm{v}}} \mathbb{1}_{\mho_{k_\circ}} \leq (9/4)\Delta_{k_\circ}^{\mathrm{v}_U}$ it follows $\delta_{k_\circ}^{\widehat{\mathrm{v}}} \mathbb{1}_{\mho_{k_\circ}} \leq \delta_{k_\circ}^{\mathrm{v}_U}(\log(9/4)/\log(3)+1)$ and hence $\operatorname{pen}_{k_\circ}^{\widehat{\mathrm{v}}} \mathbb{1}_{\mho_{k_\circ}} \leq \mathfrak{C}\Delta_{k_\circ}^{\mathrm{v}_U} \delta_{k_\circ}^{\mathrm{v}_U} k_\circ n^{-1}$, which together with $\mathbb{E}[\widehat{\sigma}_Y^2] = \sigma_Y^2$ implies

$$\mathbb{E}\left[\left\| \widehat{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X \right\|_{\mathbb{L}^2(\mathrm{v})}^2\right]$$
$$\leq 15 \left\| \mathbb{1}_{[-k_\circ,k_\circ]^C} \mathrm{M}_X \right\|_{\mathbb{L}^2(\mathrm{v})}^2 + \mathfrak{C}(\eta_U^4 \cdot \mathbb{E}[X_1^{2(c-1)}] + \sigma_Y^2) \cdot \Delta_{k_\circ}^{\mathrm{v}_U} \delta_{k_\circ}^{\mathrm{v}_U} k_\circ n^{-1}$$
$$+ \mathfrak{C}\eta_U^4 \left\| \mathrm{M}_X \left(1 \vee |\,\mathrm{M}_U\,|^2 m\right)^{-1/2} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 + \mathfrak{C} \cdot \eta_Y (1 \vee \eta_U^2 k_Y \Delta_{k_Y}^{\mathrm{v}}) \cdot n^{-1}$$
$$+ 3\mathbb{E}\left[\left\| \mathbb{1}_{[-k_n,k_n]} (\widehat{\mathrm{M}}_Y - \mathrm{M}_Y) \widehat{\mathrm{M}}_U^{\dagger} \mathbb{1}_{\mathfrak{M}} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 (\mathbb{1}_{\mho_{k_\circ}^C} + \mathbb{1}_{\mathfrak{A}^C})\right]$$
$$+ 6 \left\| \mathrm{M}_X \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{P}(\mho_{k_\circ}^C). \tag{59}$$

Since $\left\| \widehat{\mathrm{M}}_U^{\dagger} \mathbb{1}_{\mathfrak{M}} \right\|_{\mathbb{L}^\infty}^2 \leq m$ and $\left\| \mathbb{1}_{[-k_n,k_n]} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \leq 2k_n \Delta_{k_n}^{\mathrm{v}} \leq 2n\Delta_1^{\mathrm{v}}$ due to the definition (34) of k_n we obtain

$$\mathbb{E}\left[\left\| \mathbb{1}_{[-k_n,k_n]} (\widehat{\mathrm{M}}_Y - \mathrm{M}_Y) \widehat{\mathrm{M}}_U^{\dagger} \mathbb{1}_{\mathfrak{M}} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\mho_{k_\circ}^C}\right]$$
$$\leq \mathbb{E}\left[\left\| \mathbb{1}_{[-k_n,k_n]} (\widehat{\mathrm{M}}_Y - \mathrm{M}_Y) \right\|_{\mathbb{L}^2(\mathrm{v})}^2\right] m\mathbb{P}(\mho_{k_\circ}^C)$$
$$\leq \sigma_Y^2 n^{-1} \left\| \mathbb{1}_{[-k_n,k_n]} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 m\mathbb{P}(\mho_{k_\circ}^C) \leq 2\sigma_Y^2 \Delta_1^{\mathrm{v}} m\mathbb{P}(\mho_{k_\circ}^C).$$

Moreover, for each $t \in \mathbb{R}$ we have

$$\mathbb{E}\left[|\widehat{\mathrm{M}}_Y(t) - \mathrm{M}_Y(t)|^2 \mathbb{1}_{\mathfrak{A}^C}\right] \leq \mathfrak{C}n^{-1} (\mathbb{E}(Y_1^{4(c-1)})^{1/2} (\mathbb{P}(\mathfrak{A}^C)^{1/2}$$
$$\leq \mathfrak{C}n^{-1} \eta_Y^{1/2} (\mathbb{P}(\mathfrak{A}^C)^{1/2}$$

and hence by exploiting $\left\| \widehat{\mathrm{M}}_U^{\dagger} \mathbb{1}_{\mathfrak{M}} \right\|_{\mathbb{L}^\infty}^2 \leq n$ and $\left\| \mathbb{1}_{[-k_n,k_n]} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \leq 2n\Delta_1^{\mathrm{v}}$ we obtain

$$\mathbb{E}\left[\left\| \mathbb{1}_{[-k_n,k_n]} (\widehat{\mathrm{M}}_Y - \mathrm{M}_Y) \widehat{\mathrm{M}}_U^{\dagger} \mathbb{1}_{\mathfrak{M}} \right\|_{\mathbb{L}^2(\mathrm{v})}^2 \mathbb{1}_{\mathfrak{A}^C}\right] \leq \mathfrak{C}\eta_Y^{1/2} \Delta_1^{\mathrm{v}} n (\mathbb{P}(\mathfrak{A}^C)^{1/2}.$$

Since $\mathfrak{A}^C \subseteq \{|\hat\sigma_Y - \sigma_Y| > \sigma_Y/2\}$ and $\sigma_Y \geq 1$ by Markov's inequality we have $\mathbb{P}(\mathfrak{A}^C) \leq \mathfrak{C}\mathbb{E}[Y_1^{16(c-1)}]n^{-4} \leq \mathfrak{C}\eta_Y n^{-4}$. Combining the last bounds we conclude

$$\mathbb{E}\left[\left\|\mathbb{1}_{[-k_n,k_n]}(\widehat{\mathrm{M}}_Y - \mathrm{M}_Y)\widehat{\mathrm{M}}_U^{\dagger}\mathbb{1}_{\mathfrak{M}}\right\|^2_{\mathbb{L}^2(\mathrm{v})}(\mathbb{1}_{\mho^C_{k_\circ}} + \mathbb{1}_{\mathfrak{A}^C})\right] \leq 2\sigma_Y^2\Delta_1^{\mathrm{v}} m\mathbb{P}(\mho^C_{k_\circ}) + \mathfrak{C}\eta_Y\Delta_1^{\mathrm{v}} n^{-1}$$

which together with (59) and the definition of η_X implies

$$\begin{aligned}\mathbb{E}\left[\left\|\widehat{\mathrm{M}}_X^{\hat{k}} - \mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})}\right] &\leq 15\cdot\left\|\mathbb{1}_{[-k_\circ,k_\circ]^C}\mathrm{M}_X\right\|^2_{\mathbb{L}^2(\mathrm{v})} + \mathfrak{C}\cdot(\eta_U^4\eta_X + \sigma_Y^2)\cdot\Delta_{k_\circ}^{\mathrm{v}_U}\delta_{k_\circ}^{\mathrm{v}_U}k_\circ n^{-1} \\ &+ \mathfrak{C}\cdot\eta_U^4\left\|\mathrm{M}_X\left(1\vee|\mathrm{M}_U|^2 m\right)^{-1/2}\right\|^2_{\mathbb{L}^2(\mathrm{v})} + \mathfrak{C}\cdot\eta_Y(1\vee\eta_U^2 k_Y\Delta_{k_Y}^{\mathrm{v}})\cdot n^{-1} \\ &+ 6(\sigma_Y^2\Delta_1^{\mathrm{v}} + \eta_X)m\mathbb{P}(\mho^C_{k_\circ}).\end{aligned}$$

Since the last bound is valid for all $k_\circ \in [\![k_n]\!]$ we immediately obtain the claim (37), which completes the proof. □

Proof of Proposition 4.4. The proof follows along the lines of the proof of Theorem 4.1 in [26]. In order to show the claim, we need some definitions and notations. For two functions $l, u : \mathbb{R} \to \mathbb{R}$ with $l \leq u$ introduce the bracket

$$[l,u] := \{f : \mathbb{R} \to \mathbb{R} : \, l \leq f \leq u\}.$$

For a set of functions $\mathscr{G}$ and $\varepsilon \in \mathbb{R}_{>0}$ we denote by $\mathrm{N}_{[\cdot]}(\varepsilon, \mathscr{G})$ the minimum number of brackets $[l_i, u_i]$, satisfying $\mathbb{E}[(u_i(Z_1) - l_i(Z_1))^2] \leq \varepsilon^2$, that are needed to cover $\mathscr{G}$. The associated bracking entropy integral is defined as

$$\mathrm{J}_{[\cdot]}(\delta, \mathscr{G}) := \int_{(0,\delta)} \sqrt{\log \mathrm{N}_{[\cdot]}(\varepsilon, \mathscr{G})}\mathrm{d}\lambda(\varepsilon), \quad \forall \delta \in \mathbb{R}_{>0}.$$

Further, a function $\bar f$ is called an envelope of $\mathscr{G}$, if $|f| \leq \bar f$ for all $f \in \mathscr{G}$. Analogously as presented in [26], we aim to apply Lemma 19.34 and Corollary 19.35 from [29]. Hence, we start by decomposing c_m into its real and imaginary part, namely

$$\mathfrak{Re}(c_m(t)) = \frac{1}{\sqrt{m}}\sum_{j\in[\![m]\!]}\left\{Z_j^\beta\cos(\log(2\pi t Z_j)) - \mathbb{E}\left[Z_j^\beta\cos(2\pi t\log(Z_j))\right]\right\}$$

and

$$\mathfrak{Im}(c_m(t)) = \frac{1}{\sqrt{m}}\sum_{j\in[\![m]\!]}\left\{Z_j^\beta\sin(\log(2\pi t Z_j)) - \mathbb{E}\left[Z_j^\beta\sin(2\pi t\log(Z_j))\right]\right\},$$

such that $c_m(t) = \mathfrak{Re}(c_m(t)) + \iota \cdot \mathfrak{Im}(c_m(t))$ for all $t \in \mathbb{R}$. Therefore, define the following class of functions

$$\mathscr{G}_\beta := \left\{\mathbb{R}_{>0} \ni z \mapsto \bar{\mathrm{w}}(t) z^\beta \cos(2\pi t \log(z)) \in \mathbb{R} : t \in \mathbb{R}\right\} \\ \cup \left\{\mathbb{R}_{>0} \ni z \mapsto \bar{\mathrm{w}}(t) z^\beta \sin(2\pi t \log(z)) \in \mathbb{R} : t \in \mathbb{R}\right\},$$

whose envelope is given by $f(z) := z^\beta$. Now applying Lemma 19.34 of [29] and following the argumentation of the proof of Corollary 19.35 within, we conclude

$$\mathbb{E}\left[\|c_m\|_{\mathbb{L}^\infty(\bar{\mathrm{w}})}\right] \leq \mathfrak{C}\left(\sqrt{\mathbb{E}[Z_1^{2\beta}]} + \mathrm{J}_{[\cdot]}(\sqrt{\mathbb{E}[Z_1^{2\beta}]}, \mathscr{G}_\beta)\right) \leq \mathfrak{C}\left(\eta + \mathrm{J}_{[\cdot]}(\eta, \mathscr{G}_\beta)\right).$$

As $\eta \in \mathbb{R}_{>0}$, it suffices to show that the entropy integral is finite. Hence, inspired by [33], we set

$$B_\varepsilon := \inf\left\{b \in \mathbb{R}_{>0} : \mathbb{E}[Z_1^{2\beta} \mathbb{1}_{\{|\log(Z_1)| > b\}}] \leq \varepsilon^2\right\} \\ \leq \left(\frac{\mathbb{E}[Z_1^{2\beta} |\log(Z_1)|^\gamma]}{\varepsilon^2}\right)^{\frac{1}{\gamma}} \leq \left(\frac{\eta}{\varepsilon^2}\right)^{\frac{1}{\gamma}} \tag{60}$$

due to the generalised Markov inequality. Furthermore, for grid points $t_j \in \mathbb{R}$ specified below, we define

$$g_j^\pm(z) := \left(\bar{\mathrm{w}}(t_j) z^\beta \cos(2\pi t_j \log(z)) \pm \varepsilon z^\beta\right) \mathbb{1}_{[0,B_\varepsilon]}(|\log(z)|) \\ \pm \|\bar{\mathrm{w}}\|_\infty z^\beta \mathbb{1}_{(B_\varepsilon,\infty)}(|\log(z)|),$$

as well as

$$h_j^\pm(z) := \left(\bar{\mathrm{w}}(t_j) z^\beta \sin(2\pi t_j \log(z)) \pm \varepsilon z^\beta\right) \mathbb{1}_{[0,B_\varepsilon]}(|\log(z)|) \\ \pm \|\bar{\mathrm{w}}\|_\infty z^\beta \mathbb{1}_{(B_\varepsilon,\infty)}(|\log(z)|).$$

We obtain

$$\mathbb{E}\left[\left(g_j^+(Z_1) - g_j^-(Z_1)\right)^2\right] \\ \leq 4\varepsilon^2 \mathbb{E}\left[Z_1^{2\beta} \mathbb{1}_{[0,B_\varepsilon]}(|\log(Z_1)|)\right] + 4\|\bar{\mathrm{w}}\|_\infty^2 \mathbb{E}\left[Z_1^{2\beta} \mathbb{1}_{\{|\log(Z_1)| > B_\varepsilon\}}\right] \\ \leq 4\varepsilon^2 \left(\mathbb{E}[Z_1^{2\beta}] + \|\bar{\mathrm{w}}\|_\infty^2\right).$$

and analogously

$$\mathbb{E}\left[\left(h_j^+(Z_1) - h_j^-(Z_1)\right)^2\right] \leq 4\varepsilon^2 \left(\mathbb{E}[Z_1^{2\beta}] + \|\bar{\mathrm{w}}\|_\infty^2\right)$$

It remains to choose the grid points t_j in such a way that the brackets cover the set $\mathscr{G}_\beta$. Let $t \in \mathbb{R}$ be arbitrarily chosen and take some arbitrary grid point t_j. Then with the Lipschitz constant $L_{\bar{\mathrm{w}}}$ of the density function $\bar{\mathrm{w}}$, we have for

$z \in \mathbb{R}_{>0}$ evidently $|\bar{\mathrm{w}}(t)z^\beta \cos(2\pi t \log(z)) - \bar{\mathrm{w}}(t_j)z^\beta \cos(2\pi t_j \log(z))| \leq z^\beta(\bar{\mathrm{w}}(t) + \bar{\mathrm{w}}(t_j))$ and

$$\begin{aligned}|\bar{\mathrm{w}}(t)z^\beta \cos(2\pi t \log(z)) - \bar{\mathrm{w}}(t_j)z^\beta \cos(2\pi t_j \log(z))|\\ \leq |\bar{\mathrm{w}}(t) - \bar{\mathrm{w}}(t_j)| \cdot z^\beta \cdot |\cos(2\pi t \log(z))|\\ + |\bar{\mathrm{w}}(t_j)| \cdot z^\beta \cdot |\cos(2\pi t_j \log(z)) - \cos(2\pi t \log(z))|\\ \leq L_{\bar{\mathrm{w}}}|t - t_j| \cdot z^\beta + 2\pi \|\bar{\mathrm{w}}\|_\infty \cdot z^\beta \cdot |t - t_j| \cdot B_\varepsilon.\end{aligned}$$

Hence, the function $\mathbb{R}_{>0} \ni z \mapsto \bar{\mathrm{w}}(t)z^\beta \cos(2\pi t \log(z))$ is contained in the bracket $[g_j^-, g_j^+]$, if

$$\min \{|t_j - t|(L_{\bar{\mathrm{w}}} + 2\pi \|\bar{\mathrm{w}}\|_\infty B_\varepsilon), \bar{\mathrm{w}}(t) + \bar{\mathrm{w}}(t_j)\} \leq \varepsilon. \tag{61}$$

Thus, for integer $j \in [-J_\varepsilon, J_\varepsilon]$ we are choosing the grid points in the following way:

$$t_j = \frac{j\varepsilon}{(L_{\bar{\mathrm{w}}} + 2\pi \|\bar{\mathrm{w}}\|_\infty B_\varepsilon)}$$

where J_ε is the smallest integer such that t_{J_ε} is greater than or equal to

$$T_\varepsilon := \inf \left\{ t \in \mathbb{R}_{>0} : \sup_{v:|v|\geq t} \bar{\mathrm{w}}(v) \leq \varepsilon/2 \right\}$$

satisfying $\log(T_\varepsilon) = O(\varepsilon^{-\kappa})$ with $\kappa := 1/(\rho + 1/2)$. Evidently, there are at most $2J_\varepsilon + 1$ of those grid points and, hence $\mathrm{N}_{[\cdot]}(\varepsilon, \mathscr{G}) \leq 2(2J(\varepsilon) + 1)$. Keeping the bound (60) in mind we also have $\log(B_\varepsilon/\varepsilon) = O(\log(\varepsilon^{-1-2/\gamma}))$ and thus from the inequality

$$J_\varepsilon \leq T_\varepsilon(L_{\bar{\mathrm{w}}} + 2\pi \|\bar{\mathrm{w}}\|_\infty B_\varepsilon)\varepsilon^{-1} + 1$$

we obtain $\log \mathrm{N}_{[\cdot]}(\varepsilon, \mathscr{G}) = O(\log(J(\varepsilon))) = O(\varepsilon^{-\kappa} + \log(\varepsilon^{-1-2/\gamma})) = O(\varepsilon^{-\kappa})$. Since $\kappa < 2$ we conclude

$$\int_{(0,\delta)} \sqrt{\log \mathrm{N}_{[\cdot]}(\varepsilon, \mathscr{G}_\beta)} \mathrm{d}\lambda(\varepsilon) < \infty, \quad \forall \delta \in \mathbb{R}_{>0}$$

which completes the proof. ☐

Proof of Proposition 4.5. For $m \in [\![m_\circ]\!]$ we trivially have $\mathbb{P}(\mho_{k_m}^C) \leq m_\circ^2 m^{-2}$. Therefore, let $m \in \mathbb{N}_{\geq m_\circ}$. We decompose $\widehat{\mathrm{M}}_U = \widehat{\mathrm{M}}_Z + \widehat{\mathrm{M}}_U^{\mathrm{ub}}$ into a bounded part $\widehat{\mathrm{M}}_Z$ and an unbounded part $\widehat{\mathrm{M}}_U^{\mathrm{ub}}$ (similar to Remark A.1). Precisely, for $b \in \mathbb{R}_{\geq 1}$ (to be specified below) introduce the random variables

$$Z_j := U_j \mathbb{1}_{(0,b]}(U_j^{c-1}) + \mathbb{1}_{(b,\infty)}(U_j^{c-1}), \quad \forall j \in [\![m]\!]$$

and form accordingly for each $t \in \mathbb{R}$ the empirical Mellin transform

$$\widehat{\mathrm{M}}_Z(t) := \frac{1}{m} \sum_{j \in [\![m]\!]} Z_j^{c-1+\iota 2\pi t}$$

$$= \frac{1}{m} \sum_{j \in [\![m]\!]} \{U_j^{c-1} \exp(\iota 2\pi t \log(U_j)) \mathbb{1}_{(0,b]}(U_j^{c-1}) + \mathbb{1}_{(b,\infty)}(U_j^{c-1})\}.$$

Evidently, the unbounded part $\widehat{\mathrm{M}}_U^{\mathrm{ub}} := \widehat{\mathrm{M}}_U - \widehat{\mathrm{M}}_Z$ satisfies

$$\widehat{\mathrm{M}}_U^{\mathrm{ub}}(t) = \frac{1}{m} \sum_{j \in [\![m]\!]} (U_j^{c-1+\iota 2\pi t} - 1) \mathbb{1}_{(b,\infty)}(U_j^{c-1}), \quad \forall t \in \mathbb{R}.$$

Exploiting the decomposition we obtain the elementary bound

$$\begin{aligned} \mathbb{P}(\mho_{k_m}^C) \le \mathbb{P}\left(\exists t \in [-k_m, k_m] : \, |\widehat{\mathrm{M}}_Z(t) - \mathbb{E}[\widehat{\mathrm{M}}_Z(t)]| > \frac{1}{6} |\, \mathrm{M}_U(t)|\right) \\ + \mathbb{P}\left(\exists t \in [-k_m, k_m] : \, |\widehat{\mathrm{M}}_U^{\mathrm{ub}}(t) - \mathbb{E}[\widehat{\mathrm{M}}_U^{\mathrm{ub}}(t)]| > \frac{1}{6} |\, \mathrm{M}_U(t)|\right) \end{aligned} \tag{62}$$

where we estimate separately the two terms of the bound starting with the first one. Evidently, multiplying with the density function $\bar{\mathrm{w}}$ given in (39) and making use of the definition of k_m given in (41) we have

$$\begin{aligned} &\mathbb{P}\left(\exists t \in [-k_m, k_m] : \, |\widehat{\mathrm{M}}_Z(t) - \mathbb{E}[\widehat{\mathrm{M}}_Z(t)]| > \frac{1}{6} |\, \mathrm{M}_U(t)|\right) \\ &\le \mathbb{P}\left(\exists t \in [-k_m, k_m] : \, \bar{\mathrm{w}}(t) |\widehat{\mathrm{M}}_Z(t) - \mathbb{E}[\widehat{\mathrm{M}}_Z(t)]| > \tfrac{1}{6} \inf_{s \in [-k_m, k_m]} \bar{\mathrm{w}}(s) |\, \mathrm{M}_U(s)|\right) \\ &\qquad \le \mathbb{P}\left(\exists t \in [-k_m, k_m] : \, \bar{\mathrm{w}}(t) |\widehat{\mathrm{M}}_Z(t) - \mathbb{E}[\widehat{\mathrm{M}}_Z(t)]| > \tau_m m^{-1/2}\right). \end{aligned}$$

By continuity of $t \mapsto \bar{\mathrm{w}}(t)(\widehat{\mathrm{M}}_Z(t) - \mathbb{E}[\widehat{\mathrm{M}}_Z(t)])$ we obtain

$$\begin{aligned} &\mathbb{P}\left(\exists t \in [-k_m, k_m] : \, \bar{\mathrm{w}}(t) |\widehat{\mathrm{M}}_Z(t) - \mathbb{E}[\widehat{\mathrm{M}}_Z(t)]| > \tau_m m^{-1/2}\right) \\ &\qquad \le \mathbb{P}\left(\sup_{t \in \mathcal{T}_m} |\overline{\nu}_t| > \tau_m m^{-1/2}\right) \end{aligned}$$

setting $\mathcal{T}_m := [-k_m, k_m] \cap \mathbb{Q}$ and $\overline{\nu}_t = m^{-1} \sum_{i \in [\![m]\!]} \{\nu_t(Z_i) - \mathbb{E}\,(\nu_t(Z_i))\}$ with $\nu_t(z) := z^{c-1+\iota 2\pi t} \bar{\mathrm{w}}(t)$ for $z \in \mathbb{R}$. Observe that each Z_i takes values in $\mathcal{Z} := \{z \in \mathbb{R}_{>0} : z^{c-1} \in (0, b]\}$ only. Since $\mathcal{T}_m \subset \mathbb{R}$ is countable we eventually apply Talagrand's inequality given in Lemma C.1 where we need to determine the three quantities ψ, Ψ and τ. Consider ψ first. Evidently, since $\|\bar{\mathrm{w}}\|_\infty = 1$ we have

$$\sup_{t \in \mathbb{R}} \left\{ \sup_{z \in \mathcal{Z}} |\nu_t(z)| \right\} \le b \, \|\bar{\mathrm{w}}\|_\infty = b =: \psi.$$

Consider next τ. Making use of

$$\mathbb{E}[Z_1^{2(c-1)}] = \mathbb{E}[U_1^{2(c-1)} \mathbb{1}_{(0,b]}(U_1^{c-1}) + \mathbb{1}_{(b,\infty)}(U_1^{c-1})] \le \mathbb{E}[U_1^{2(c-1)}] \le \eta_U^2$$

(keep in mind that $b \in \mathbb{R}_{\geq 1}$ and $\|\bar{\mathrm{w}}\|_\infty^2 = 1$) for each $t \in \mathbb{R}$ we have

$$\mathbb{E}\big[|\nu_t(Z_1)|^2\big] \leq \|\bar{\mathrm{w}}\|_\infty^2\, \mathbb{E}\big[Z_1^{2(c-1)}\big] \leq \eta_U^2 =: \tau.$$

Finally, consider Ψ. Recalling the normalised Mellin function process c_m defined in (38) (with $\beta = c-1$) for all $t \in \mathbb{R}$ we have $m^{-1/2}\bar{\mathrm{w}}(t)c_m(t) = \overline{\nu}_t$. Since $(\mathbb{E}[Z_1^{2(c-1)}])^{1/2} \leq (\mathbb{E}[U_1^{2(c-1)}])^{1/2} \leq \eta_U$ and

$$\mathbb{E}[Z_1^{2(c-1)}|\log(Z_1)|^\gamma] = \mathbb{E}[U_1^{2(c-1)}|\log(U_1)|^\gamma]\mathbb{1}_{(0,b]}(U_j^{c-1})] \\ \leq \mathbb{E}[U_1^{2(c-1)}|\log(U_1)|^\gamma] \leq \eta_U$$

from Proposition 4.4 it follows

$$\mathbb{E}\left[\sup_{t\in\mathcal{T}_m}|\overline{\nu}_t|\right] = m^{-1/2}\mathbb{E}\left[\sup_{t\in\mathcal{T}_m}\bar{\mathrm{w}}(t)|c_m(t)|\right] \\ \leq m^{-1/2}\mathbb{E}\left[\|c_m\|_{\mathbb{L}^\infty(\bar{\mathrm{w}})}\right] \leq m^{-1/2}\mathfrak{C}(\eta_U,\rho) =: \Psi.$$

Due to Lemma C.1 with $\kappa_m := (\tau_m - 2\mathfrak{C}(\eta_U,\rho))m^{-1/2} \in \mathbb{R}_{>0}$ we obtain

$$\mathbb{P}\left(\exists t \in [-k_m,k_m] : |\widehat{\mathrm{M}}_Z(t) - \mathbb{E}[\widehat{\mathrm{M}}_Z(t)]| > \frac{1}{6}|\,\mathrm{M}_U(t)|\right) \\ \leq \mathbb{P}\left(\sup_{t\in\mathcal{T}_m}|\overline{\nu}_t| > \tau_m m^{-1/2}\right) = \mathbb{P}\left(\sup_{t\in\mathcal{T}_m}|\overline{\nu}_t| > m^{-1/2}2\mathfrak{C}(\eta_U,\rho) + \kappa_m\right) \\ \leq 3\exp\left(-\mathfrak{C}_{\mathrm{tal}}\big(\frac{(\tau_m - 2\mathfrak{C}(\eta_U,\rho))^2}{\eta_U^2} \wedge \frac{m^{1/2}(\tau_m - 2\mathfrak{C}(\eta_U,\rho))}{b}\big)\right).$$

Setting $b = 2m^{1/2}\eta_U^2/(\tau_m - 2\mathfrak{C}(\eta_U,\rho)) = m^{1/2}(\log m)^{-1/2}\eta_U\mathfrak{C}_{\mathrm{tal}}^{1/2} \in \mathbb{R}_{\geq 1}$ (keep (40) in mind) it follows

$$\mathbb{P}\left(\exists t \in [-k_m,k_m] : |\widehat{\mathrm{M}}_Z(t) - \mathbb{E}[\widehat{\mathrm{M}}_Z(t)]| > \frac{1}{6}|\,\mathrm{M}_U(t)|\right) \\ \leq 3\exp\left(-\mathfrak{C}_{\mathrm{tal}}\frac{(\tau_m - 2\mathfrak{C}(\eta_U,\rho))^2}{2\eta_U^2}\right) = 3m^{-2}. \quad (63)$$

Consider the second term on the right hand side of (62). For each $t \in \mathbb{R}$ (keep $b \in \mathbb{R}_{\geq 1}$ in mind) we have

$$|\widehat{\mathrm{M}}_U^{\mathrm{ub}}(t)]| \leq \frac{1}{m}\sum_{j\in[\![m]\!]}|U_j^{c-1+\iota 2\pi t} - 1|\mathbb{1}_{(b,\infty)}(U_j^{c-1}) \\ \leq \frac{1}{m}\sum_{j\in[\![m]\!]}(U_j^{c-1}+1)\mathbb{1}_{(b,\infty)}(U_j^{c-1}) \leq \frac{2}{m}\sum_{j\in[\![m]\!]}U_j^{c-1}\mathbb{1}_{(b,\infty)}(U_j^{c-1})$$

and hence $|\mathbb{E}[\widehat{\mathrm{M}}_U^{\mathrm{ub}}(t)]| \le 2b^{-1}\mathbb{E}[U_1^{2(c-1)}] \le 2b^{-1}\eta_U^2 = m^{-1/2}(\tau_m - 2\mathfrak{C}(\eta_U, \rho))$. Multiplying the density function $\bar{\mathrm{w}} : \mathbb{R} \to (0, 1]$ given in (39) and making use of the definition (41) of k_m it follows (with $\mathfrak{C}(\eta_U, \rho) \in \mathbb{R}_{\ge 1}$)

$$\begin{aligned}
&\mathbb{P}\bigg(\exists t \in [-k_m, k_m] : |\widehat{\mathrm{M}}_U^{\mathrm{ub}}(t) - \mathbb{E}[\widehat{\mathrm{M}}_U^{\mathrm{ub}}(t)]| > \frac{1}{6}|\mathrm{M}_U(t)|\bigg) \\
&\quad \le \mathbb{P}\bigg(\exists t \in [-k_m, k_m] : |\widehat{\mathrm{M}}_U^{\mathrm{ub}}(t)| > \frac{1}{6}\bar{\mathrm{w}}(t)|\mathrm{M}_U(t)| - |\mathbb{E}[\widehat{\mathrm{M}}_U^{\mathrm{ub}}(t)]|\bigg) \\
&\quad \le \mathbb{P}\bigg(\frac{2}{m}\sum_{j \in [\![m]\!]} U_j^{c-1}\mathbb{1}_{(b,\infty)}(U_j^{c-1}) > \frac{1}{6}\inf_{v \in [-k_m, k_m]} \bar{\mathrm{w}}(v)|\mathrm{M}_U(v)| \\
&\qquad\qquad - m^{-1/2}(\tau_m - 2\mathfrak{C}(\eta_U, \rho))\bigg) \\
&\quad \le \mathbb{P}\bigg(\sum_{j \in [\![m]\!]} U_j^{c-1}\mathbb{1}_{(b,\infty)}(U_j^{c-1}) > \mathfrak{C}(\eta_U, \rho)m^{1/2}\bigg) \\
&\quad \le m^{1/2}\mathbb{E}\left[U_1^{c-1}\mathbb{1}_{(b,\infty)}(U_1^{c-1})\right] \le m^{1/2}b^{-6}\mathbb{E}[U_1^{7(c-1)}] \\
&\quad \le \mathfrak{C}_{\mathrm{tal}}^{-3}\eta_U m^{-1/2}(\log m)^3 m^{-2} \le (6/e)^3\mathfrak{C}_{\mathrm{tal}}^{-3}\eta_U m^{-2} \le 11\mathfrak{C}_{\mathrm{tal}}^{-3}\eta_U m^{-2}.
\end{aligned}$$

Combining the last upper bound, the upper bound (63) and the decomposition (62) we obtain the claim which completes the proof. □

Acknowledgments

The authors would like to thank the anonymous referees, an Associate Editor and the Editor for their constructive comments that improved the quality of this paper.

Funding

This work is funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC-2181/1-39090098 (the Heidelberg STRUCTURES Cluster of Excellence).

References

[1] Andersen, K. E. and Hansen, M. B. (2001). Multiplicative censoring: density estimation by a series expansion approach. *Journal of Statistical Planning and Inference* **98** 137–155. https://doi.org/10.1016/S0378-3758(00)00237-8 MR1860231.

[2] 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** 2109–2131. https://doi.org/10.1214/009053605000000372 MR2211081.

[3] Barron, A., Birgé, L. and Massart, P. (1999). Risk bounds for model selection via penalization. *Probability Theory and Related Fields* **113** 301-413. https://doi.org/10.1007/s004400050210. MR1679028

[4] Belomestny, D., Comte, F. and Genon-Catalot, V. (2016). Nonparametric Laguerre estimation in the multiplicative censoring model. *Electronic Journal of Statistics* **10** 3114–3152. https://doi.org/10.1214/16-EJS1203 MR3571964.

[5] 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** 36–67. https://doi.org/10.1214/18-AIHP954 MR4058981.

[6] Belomestny, D. and Goldenshluger, A. (2021). Density deconvolution under general assumptions on the distribution of measurement errors. *The Annals of Statistics* **49**. https://doi.org/10.1214/20-aos1969. MR4255101

[7] Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli* **4** 329-375. https://doi.org/10.2307/3318720. MR1653272

[8] Brenner Miguel, S. (2022). Anisotropic spectral cut-off estimation under multiplicative measurement errors. *Journal of Multivariate Analysis* **190** 104990. https://doi.org/10.1016/j.jmva.2022.104990. MR4396579

[9] Brenner Miguel, S. (2023). The Mellin transform in nonparametric statistics, PhD thesis, Ruprecht-Karls-Universität Heidelberg, 2023. https://doi.org/10.11588/heidok.00029255.

[10] 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** 3551 – 3573. https://doi.org/10.1214/21-EJS1870. MR4298979

[11] Brenner Miguel, S., Comte, F. and Johannes, J. (2023). Linear functional estimation under multiplicative measurement error. *Bernoulli* **29** 2247 – 2271. https://doi.org/10.3150/22-BEJ1540. MR4580915

[12] Brenner Miguel, S. and Phandoidaen, N. (2022). Multiplicative deconvolution in survival analysis under dependency. *Statistics* **56** 297-328. https://doi.org/10.1080/02331888.2022.2058503. MR4415439

[13] Brunel, E., Comte, F. and Genon-Catalot, V. (2016). Nonparametric density and survival function estimation in the multiplicative censoring model. *TEST* **25** 570–590. https://doi.org/10.1007/s11749-016-0479-1 MR3531843.

[14] Comte, F. (2017). *Nonparametric Estimation*. Spartacus-Idh.

[15] Comte, F. and Dion, C. (2016). Nonparametric estimation in a multiplicative censoring model with symmetric noise. *Journal of Nonparametric Statistics* **28** 768-801. https://doi.org/10.1080/10485252.2016.

1225737. MR3555457

[16] Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **73** 601–627. MR2853732

[17] Comte, F., Rozenholc, Y. and Taupin, M. L. (2006). Penalized contrast estimator for density deconvolution. *Canadian Journal of Statistics* **37**. MR2328553

[18] Efromovich, S. (1999). *Nonparametric curve estimation. Methods, theory, and applications*. *Springer Ser. Stat.* New York, NY: Springer. https://doi.org/10.1007/b97679. MR1705298

[19] Engl, H. W., Hanke, M. and Neubauer, A. (1996). *Regularization of inverse problems*. *Mathematics and its Applications* **375**. Kluwer Academic Publishers Group, Dordrecht. MR1408680.

[20] Johannes, J. (2009). Deconvolution with unknown error distribution. *The Annals of Statistics* **37** 2301 – 2323. https://doi.org/10.1214/08-AOS652. MR2543693

[21] Johannes, J. and Schwarz, M. (2013). Adaptive circular deconvolution by model selection under unknown error distribution. *Bernoulli* **19** 1576 – 1611. https://doi.org/10.3150/12-BEJ422. MR3129026

[22] Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *The Annals of Probability* **33** 1060 – 1077. https://doi.org/10.1214/009117905000000044. MR2135312

[23] Massart, P. (2007). *Concentration inequalities and model selection. Ecole d'Eté de Probabilités de Saint-Flour XXXIII – 2003.* Lecture Notes in Mathematics 1896. Berlin: Springer. https://doi.org/10.1007/978-3-540-48503-2. MR2319879

[24] Meister, A. (2009). *Deconvolution problems in nonparametric statistics*. *Lecture Notes in Statistics* **193**. Springer-Verlag, Berlin. https://doi.org/10.1007/978-3-540-87557-4 MR2768576.

[25] Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics* **7** 307 – 330. MR1460203

[26] Neumann, M. H. and Reiss, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli* **15** 223 – 248. https://doi.org/10.3150/08-bej148. MR2546805

[27] Petrov, V. V. (1995). *Limit theorems of probability theory. Sequences of independent random variables.*, 4. ed. *Oxford Studies in Probability.* Clarendon Press., Oxford. MR1353441

[28] Talagrand, M. (1996). New concentration inequalities in product spaces. *Inventiones mathematicae* **126** 505–563. https://doi.org/10.1007/s002220050108. MR1419006

[29] van der Vaart, A. W. (1998). *Asymptotic statistics.* Cambridge University Press. MR1652247

[30] van Es, B., Klaassen, C. A. J. and Oudshoorn, K. (2000). Survival analysis under cross-sectional sampling: length bias and multiplicative

censoring. **91** 295–312. Prague Workshop on Perspectives in Modern Statistical Inference: Parametrics, Semi-parametrics, Non-parametrics (1998). https://doi.org/10.1016/S0378-3758(00)00183-X MR1814785.

[31] Vardi, Y. (1989). Multiplicative Censoring, Renewal Processes, Deconvolution and Decreasing Density: Nonparametric Estimation. *Biometrika* **76** 751–761. MR1041420

[32] Vardi, Y. and Zhang, C.-H. (1992). Large sample study of empirical distributions in a random-multiplicative censoring model. *The Annals of Statistics* **20** 1022–1039. https://doi.org/10.1214/aos/1176348668 MR1165604.

[33] Yukich, J. E. (1985). Weak convergence of the empirical characteristic function. *Proceedings of the American Mathematical Society* **95** 470–473. https://doi.org/10.2307/2045821. MR0806089