Universität Stuttgart

Fachbereich Mathematik

Smoothing spline regression estimates for randomly right censored data

Jürgen Dippon, Stefan Winter

Preprint 2006/006

Universität Stuttgart
Fachbereich Mathematik

Smoothing spline regression estimates for randomly right censored data

Jürgen Dippon, Stefan Winter

Preprint 2006/006

Fachbereich Mathematik
Fakultät Mathematik und Physik
Universität Stuttgart
Pfaffenwaldring 57
D-70 569 Stuttgart

E-Mail: preprints@mathematik.uni-stuttgart.de
WWW: http://www.mathematik/uni-stuttgart.de/preprints

ISSN **1613-8309**

LaTeX-Style: Winfried Geis, Thomas Merkle

Smoothing spline regression estimates for randomly right censored data

Jürgen Dippon and Stefan Winter

Fachbereich Mathematik, Universität Stuttgart, Pfaffenwaldring 57, D-70569 Stuttgart, Germany, email: {dippon, wintersn}@mathematik.uni-stuttgart.de

September 24, 2006

Abstract

Let $X \in \mathbb{R}^d$ be a random vector, Y a non-negative and bounded random variable, and C a right censoring random variable operating on Y, which is independent of (X, Y). Given a sample of the distribution of $\big(X, \min\{Y, C\}, I_{[Y<C]}\big)$, our goal is to construct estimates of the regression function $m(x) = \mathbf{E}[\,Y \mid X = x\,]$.

We prove that suitable defined smoothing spline estimates are consistent with respect to the $\mathcal{L}_2$ error and achieve the optimal rate of convergence up to a logarithmic factor.

Key words and phrases: censored data, regression estimate, universal consistency, rate of convergence, penalized least squares estimates, smoothing splines.

1 Introduction

1.1 Nonparametric regression analysis

Let (X, Y), (X_1, Y_1), (X_2, Y_2),..., $(X_n, Y_n) \in \mathbb{R}^d \times \mathbb{R}$ be independent identically distributed (i.i.d.) random vectors with $\mathbf{E}Y^2 < \infty$. No assumptions are made on the distribution functions of the coordinates of X: Some of them may be continuous, others may be step functions or a composition of these two types of distribution functions.

In *regression analysis* one wants to estimate Y after having observed X, i.e., one wishes to determine a function f such that $f(X)$ is a "good" approximation of Y. Here, we measure the "distance" between $f(X)$ and Y by the $\mathcal{L}_2$ *risk* of f,

$$\mathbf{E}\left[\,|f(X) - Y|^2\,\right], \tag{1.1}$$

which we now want to minimize. It is well known that the $\mathcal{L}_2$ risk of every measurable function f is the sum of the $\mathcal{L}_2$ risk of the *regression function* $m : \mathbb{R}^d \to \mathbb{R} : x \mapsto \mathbf{E}[\,Y \mid X = x\,]$ and the $\mathcal{L}_2$ *error* :

$$\mathbf{E}\left[\,|f(X) - Y|^2\,\right] = \mathbf{E}\left[\,|m(X) - Y|^2\,\right] + \int_{\mathbb{R}^d} |f(x) - m(x)|^2 \mu(dx). \tag{1.2}$$

Here μ denotes the distribution of X. Since the $\mathcal{L}_2$ error is always non-negative, (1.2) implies that the regression function m is the optimal predictor of Y in view of the minimization of the $\mathcal{L}_2$ risk:

$$\mathbf{E}\left[|m(X)-Y|^2\right] = \min_{\substack{f:\mathbb{R}^d\to\mathbb{R},\\ f\text{measurable}}} \mathbf{E}\left[|f(X)-Y|^2\right]. \tag{1.3}$$

In practical applications, the distribution of (X, Y) and hence also m are usually unknown. But it is often possible to observe a sample $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ of this distribution, and one can construct estimates

$$m_n(\cdot) := m_n(\cdot, (X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)) : \mathbb{R}^d \to \mathbb{R}$$

of the regression function.

It follows from (1.2) that such an estimate m_n is a good approximation of Y in the sense that the $\mathcal{L}_2$ risk of m_n is close to the optimal value $\mathbf{E}\left[|m(X)-Y|^2\right]$ if and only if the $\mathcal{L}_2$ error $\int_{\mathbb{R}^d} |m_n(x)-m(x)|^2\mu(dx)$ is small. Consequently, the error caused by using an estimate m_n instead of m will be measured by the $\mathcal{L}_2$ error.

Definition 1.1 (Consistency) *A sequence of measurable regression estimates* $(m_n)_{n\in\mathbb{N}}$ *is called* **strongly universally consistent** *if*

$$\int_{\mathbb{R}^d} |m_n(x)-m(x)|^2\mu(dx) \to 0 \quad (n\to\infty)$$

almost surely (a.s.) for all distributions of (X, Y) *with* $\mathbf{E}Y^2 < \infty$.

1.2 Regression estimates for right censored data

Right censoring occurs whenever with non-zero probability only a lower bound on a random variable Y of interest is known. Typical examples for Y would be lifetimes of patients in a medical study or products in quality control. For a patient (or a product) not failing before leaving the study, one doesn't observe a realization y of Y, but only a lower bound c of y. In the following, we will assume *random censoring* which means that the observed censoring time c for a patient is not fixed in advance, but can be interpreted as a realization of a *right censoring variable* C operating on Y. Examples of random censoring include dropouts in medical studies or deaths unrelated to the studied causes. The regression estimation problem we are dealing with can now be formulated as follows: Let $(X, Y, C), (X_1, Y_1, C_1), (X_2, Y_2, C_2), \ldots, (X_n, Y_n, C_n)$ be i.i.d. $\mathbb{R}^d \times \mathbb{R}_+ \times \mathbb{R}_+$ – valued random vectors. In practice, beside the realizations of the covariables X_i, all we observe are realizations of the minima of Y_i and C_i $(i = 1, \ldots, n)$, and we have the information whether a censored observation has occurred or not. Set therefore $Z := \min\{Y, C\}$, $\delta := I_{[Y<C]}$, $Z_i := \min\{Y_i, C_i\}$, and $\delta_i = I_{[Y_i<C_i]}$. Here, for $a_1, a_2 \in \mathbb{R}$,

$$I_{[a_1<a_2]} := \begin{cases} 1 & \text{if} \quad a_1 < a_2 \\ \\ 0 & \text{if} \quad a_1 \geq a_2. \end{cases}$$

The problem is now to estimate the regression function from the data

$$\mathcal{D}_n := \{(X_1, Z_1, \delta_1), \ldots, (X_n, Z_n, \delta_n)\}.$$

1.3 Results from regression analysis of censored data

There are basically two methods to determine the functional interrelationship between covariates and censored response: regression based approaches, on which we focus in this article, and hazard risk approaches, which include classical Cox regression as well as extensions to nonparametric models. Details regarding the latter one can, e.g., be found in the books of Andersen, Borgan, Gill and Keiding (1993), Fleming and Harrington (1991) or Cox and Oakes (1984) and in the works of Dippon (2004), Huang and Stone(1998), Kooperberg, Stone and Troung (1995a, 1995b), and the literature cited therein.

Concerning regression based approaches, Buckley and James (1979) introduced an estimator of a linear regression function, whose consistency was investigated by James and Smith (1984). For a slight modification of this estimate, Ritov (1990) and Lai and Ying (1991) established the asymptotic normality. Other estimates for the linear regression model are due to Leurgans (1987), Koul, Sousarla, and Van Ryzin (1981), and Miller (1976).

Without conditions on the structure of the regression function or regularity assumptions on the distribution of the design, Zheng (1988) showed that suitable defined nearest neighbor estimates for censored regression are strongly pointwise consistent. He required that (X, Y) and C are independent. In the same setting, strong consistency of suitable defined partitioning estimates with respect to the $\mathcal{L}_2$ error was proven by Carbonez (1992). A survey of corresponding results for further nonparametric estimates is given in Pintér (2001). Beyond, in the more general model that Y and C are conditionally independent given X, Pintér (2001) showed that one can use a nonparametric estimate of the conditional survival function, introduced by Beran (1981), to construct suitable defined local averaging estimates which are strongly consistent with respect to the $\mathcal{L}_2$ error.

We are especially interested in the rate of convergence of nonparametric regression estimates for censored data. For the complete data model, an important theoretical breakthrough is due to Stone (1982). He showed that for (p, B)-smooth regression functions and d-dimensional covariates, the optimal global rate of convergence of nonparametric estimates is given by $n^{-\frac{2p}{2p+d}}$ (cf. Remark 3.3). Results for censored regression based on a hazard risk model can, e.g., be found in Huang and Stone (1998), Kooperberg, Stone, and Troung (1995b), and Zucker and Karr (1990).

However, little is known about the rate of convergence regarding regression based approaches for the analysis of censored data. Under regularity conditions on the distribution of X (in particular that X has a density with respect to the Lebesgue–Borel measure), Fan and Gijbels (1994) showed that suitable defined local polynomial estimates achieve pointwise the optimal rate of convergence. In the presence of right censoring and possible left truncation, Park (1999) proved that for (p, B)-smooth regression functions suitable defined weighted least squares estimates reach the optimal rate of convergence, if X has a bounded marginal density with respect to the Lebesgue–Borel measure. However, these estimates are not calculable, since they depend on p, which is unknown in a statistical application. We show that it is possible to define nonparametric regression estimates (in particular smoothing spline estimates) for censored data which achieve the optimal rate of convergence up to a logarithmic factor without assuming any regularity condition on the distribution of X (besides X is bounded), and that this result even holds for adaptive estimates, i.e., if we choose the parameters of our estimates by a completely data driven method.

1.4 Regularity assumptions

This sequel presents the regularity conditions on the underlying distributions, which we require in order to generalize known bounds on the $\mathcal{L}_2$ error of our estimates from nonparametric regression with random design on censored regression. Throughout our paper, we will use the following notation: Let $F(t) := \mathbf{P}[\,Y > t\,]$ and $G(t) := \mathbf{P}[\,C > t\,]$ $(t \in \mathbb{R})$ be the survival functions of the uncensored and censoring times, respectively. Furthermore, set $\tau_F := \sup\{t \in \mathbb{R} : F(t) > 0\}$ and $\tau_G := \sup\{t \in \mathbb{R} : G(t) > 0\}$. Our regularity assumptions can now be stated as follows:

(RA1) $X \in [0,1]^d$ a.s.

(RA2) There exists a constant $L \in [0, \infty)$ such that $0 \leq Y \leq L$ a.s., $C \geq 0$ a.s., and $\mathbf{P}[\,C > L\,] > 0$.

(RA3) C and (X, Y) are independent

(RA4) G is continuous.

To be able to give upper bounds on the covering numbers in the proof of Theorems 3.2 and 3.3, we require that X and Y are bounded in absolute value with probability one, w.l.o.g. $X \in [0,1]^d$ a.s. in **(RA1)** and $Y \in [0, L]$ a.s. in **(RA2)**, respectively (note that only for the sake of convenience we have chosen $Y, C \geq 0$ a.s. **(RA2)**). Boundedness of X is a common assumption in the analysis of the rate of convergence and is, as the boundedness of Y, not a serious constraint in a statistical application. **(RA2)** implies $Y \leq \tau_F < \infty$ a.s., but since τ_F (and τ_G) is unknown in a statistical application, we define our estimate with a more general and known upper bound L. Once this bound is determined, one can make a more or less rough estimate of τ_F and τ_G, as we know from **(RA2)** that $\tau_F \leq L < \tau_G$. Furthermore we want to stress that in **(RA2)** $\mathbf{P}[C > L] = 1$ is allowed. Therefore our main results presented in Section 3 are still valid if censoring does not occur. They can be regarded as generalizations of results for multivariate smoothing splines in usual nonparametric regression with random design (vide Kohler and Krzyżak (2001) and Kohler, Krzyżak and Schäfer (2002)) to censored regression. Assumption **(RA3)** is used to simplify the mathematical problem. It is realistic whenever the mechanism of censoring is independent of the covariables under study. Of course there exist applications where this is not satisfied, but without assumption **(RA3)** the analysis of the rate of convergence, which is the main aim of this article, seems to be much more difficult. Assumption **(RA4)** is used to simplify the presentation of our main results and their proofs. Vide Remark 3.2 and 3.6 in Section 3 for details.

1.5 Discussion of the main results

The multivariate smoothing spline estimates considered in this article are defined by:

1. transforming the censored data to virtually uncensored data as in Fan and Gijbels (1994, 1996)
2. minimizing the sum of empirical $\mathcal{L}_2$ error and a penalty term over the Sobolev space $W_k([0,1])^d$ (for the definition of $W_k([0,1])^d$, see Definition 2.1), where $k \in \mathbb{N}$ is a parameter of the estimate

We show that if **(RA1)** – **(RA4)** and an additional assumption, which controls the heaviness of the censoring near τ_F (see (3.1)), hold, these estimates achieve for smooth regression functions (i.e., $m \in W_p([0,1])^d$ for some $p \in \mathbb{N}$ with $2p > d$) the optimal global rate of convergence up to some logarithmic factor, and that this result still holds for estimates for which the parameters are chosen in a total data-dependent way (i.e., for estimates which do not depend on the smoothness of m). Furthermore, we prove that the estimates are strongly consistent even if the regression function is not smooth.

1.6 Outline

In Section 2, the smoothing spline estimates for randomly right censored data are defined. The main results are presented in Section 3 and proven in Section 6. Generalizations of the theorems in Section 3 to regression estimation with additional measurement errors in the dependent variable are given in Section 4. Section 5 contains the proofs of the results in Section 4, while auxiliary results are shown in Appendix A and B.

2 Definition of the estimate

2.1 Multivariate smoothing spline estimates (MSSE) for uncensored data

Let

$$\mathcal{D}_n := \{(X_1, Y_1), \ldots, (X_n, Y_n)\} \tag{2.1}$$

be a i.i.d. sample of the $\mathbb{R}^d \times \mathbb{R}$-valued random vector (X, Y) with $\mathbf{E}Y^2 < \infty$. Since the regression function minimizes the $\mathcal{L}_2$ risk (cf.(1.3)), a natural estimate of m can be obtained by minimizing an estimate of the $\mathcal{L}_2$ risk, the empirical $\mathcal{L}_2$ risk,

$$\frac{1}{n} \sum_{i=1}^{n} |f(X_i) - Y_i|^2. \tag{2.2}$$

But if one would minimize (2.2) over all (measurable) functions, this would lead to a function which interpolates the data (at least if the $X_1, \ldots, X_n$ are all distinct). There are basically two different strategies to avoid this: For least squares estimates one minimizes the empirical $\mathcal{L}_2$ risk over some suitable chosen class of functions which depends on the sample size n. For penalized least squares estimates or smoothing spline estimates one minimizes the sum of the empirical $\mathcal{L}_2$ risk and a penalty term which penalizes the roughness of a function, over basically all functions.

Definition 2.1 (Multivariate smoothing spline estimates (MSSE)) *Let $d, k \in \mathbb{N}$ with $2k > d$, $X \in [0,1]^d$ a.s., $\mathcal{D}_n$ be given by (2.1), and denote by $W_k([0,1]^d)$ the Sobolev space*

$$\left\{ f : \frac{\partial^\kappa f}{\partial x_1^{\kappa_1} \ldots \partial x_d^{\kappa_d}} \in \mathcal{L}_2([0,1]^d) \ \forall \kappa_1, \ldots, \kappa_d \in \mathbb{N}_0, \sum_{i=1}^{d} \kappa_i = \kappa \leq k \right\}. \tag{2.3}$$

The **multivariate smoothing spline estimate (MSSE)** $\tilde{m}_{n,(k,\lambda_n)}$ *is defined by*

$$\tilde{m}_{n,(k,\lambda_n)}(\cdot, \mathcal{D}_n) := \underset{f \in W_k([0,1]^d)}{\arg\min} \left(\frac{1}{n} \sum_{i=1}^{n} |f(X_i) - Y_i|^2 + \lambda_n J_k^2(f) \right) \tag{2.4}$$

with the parameter $\lambda_n > 0$ and the penalty term

$$J_k^2(f) := \sum_{\substack{\kappa_1, \ldots, \kappa_d \in \mathbb{N}_0: \\ \kappa_1 + \ldots + \kappa_d = k}} \frac{k!}{\kappa_1! \cdot \ldots \cdot \kappa_d!} \int_{[0,1]^d} \left| \frac{\partial^k f}{\partial x_1^{\kappa_1} \ldots \partial x_d^{\kappa_d}}(x) \right|^2 dx. \tag{2.5}$$

The condition $2k > d$ implies that the functions in $W_k([0,1]^d)$ are continuous and hence the evaluation of a function at a point is well defined.

Note that in Definition 2.1 of the estimate we do not require that the minimizer is unique. Duchon (1976) and (under some additional assumptions) Wahba (1990) showed that a function of the form

$$\sum_{i=1}^{n} a_{1,i} R(\|x - X_i\|) + \sum_{j=1}^{N} a_{2,j} \Psi_j(x) \quad (x \in \mathbb{R}^d)$$

achieves the minimum in (2.4), where

$$R : \mathbb{R}_+ \to \mathbb{R} : t \mapsto \begin{cases} t^{2k-d} \ln t & \text{if} \quad 2k - d \quad \text{is even} \\ \\ t^{2k-d} & \text{if} \quad 2k - d \quad \text{is odd,} \end{cases}$$

$\Psi_1, \ldots, \Psi_N$ are all monomials $x_1^{\kappa_1} \cdot \ldots \cdot x_d^{\kappa_d}$ of total degree $\sum_{i=1}^{d} \kappa_i \leq k - 1$, and $\|x\|$ denotes the Euclidean norm of $x \in \mathbb{R}^d$. Furthermore, Duchon and Wahba showed that the coefficients $a_{1,1}, \ldots, a_{1,n}, a_{2,1}, \ldots, a_{2,N} \in \mathbb{R}$ can be computed by solving a linear system of equations.

2.2 MSSE for randomly right censored data

Throughout this section we assume that **(RA1)** – **(RA3)** hold. To define regression estimates for censored data, we first transform the data according to Fan and Gijbels (1994, 1996). Based on these data, estimates are defined as in usual nonparametric regression (cf. Section 2.1). Therefore the transformation has to be defined in such a way that the regression functions of both the transformed data and the censored data are identical.

This so called *censoring unbiased transformation* has been investigated by many authors, for example Buckley and James (1979), Koul, Susarla, and van Ryzin (1981), Leurgans (1987), Zheng (1987), or Fan and Gijbels (1994, 1996).

To be more precise, a censored datum point (X, Z, δ) will be replaced by (X, Y^*) where

$$Y^* := \delta\, \Phi_1(Z) + (1 - \delta)\, \Phi_2(Z) = \begin{cases} \Phi_1(Y) & \text{if} \quad Y < C \\ \\ \Phi_2(C) & \text{if} \quad Y \geq C \end{cases} \tag{2.6}$$

and the transformation functions $\Phi_1(\cdot)$ and $\Phi_2(\cdot)$ are chosen such that

$$\mathbf{E}\,[\, Y^* \,|\, X \,] = m(X) = \mathbf{E}\,[\, Y \,|\, X \,]\,. \tag{2.7}$$

A special family of transformations satisfying (2.7) is given by the following two functions (see Fan and Gijbels (1994, 1996)):

$$\Phi_1(Z) \quad := \quad (1 + \alpha) \int_0^Z \frac{dt}{G(t)} - \alpha \frac{Z}{G(Z)},$$

$$\Phi_2(Z) \quad := \quad (1+\alpha)\int_0^Z \frac{dt}{G(t)}. \tag{2.8}$$

Here $\alpha \in \mathbb{R}$ is the parameter of the transformation. One could, e.g., choose α such that $Y^* \geq 0$ a.s. (corresponding to $Y \geq 0$ a.s.), which is for example fulfilled for $\alpha = 0$ (Leurgans (1987)) or $\alpha = -1$ (Koul, Susarla, and van Ryzin (1981)). Fan and Gijbels (1994, 1996) suggested a data-dependent choice of the parameter:

$$\alpha = \min_{\substack{i=1,\ldots,n:\\ \delta_i=1}} \frac{\int_0^{Z_i} \frac{dt}{G(t)} - Z_i}{\frac{Z_i}{G(Z_i)} - \int_0^{Z_i} \frac{dt}{G(t)}}.$$

Note that our main results presented in Section 3 are valid for any (fixed) $\alpha \in \mathbb{R}$. According to (2.6) and (2.8), we replace (X, Z, δ) by (X, Y^*) with

$$Y^* = (1+\alpha)\int_0^Z \frac{dt}{G(t)} - \alpha\frac{\delta Z}{G(Z)} \tag{2.9}$$

$\left(\frac{0}{0} := 0\right)$ and, for all $i = 1, \ldots, n$, (X_i, Z_i, δ_i) by $(X_i, Y_i{}^*)$, where

$$Y_i{}^* = (1+\alpha)\int_0^{Z_i} \frac{dt}{G(t)} - \alpha\frac{\delta_i Z_i}{G(Z_i)}. \tag{2.10}$$

For Y^* chosen as in (2.9), it is not hard to check that (2.7) holds for all $\alpha \in \mathbb{R}$. From **(RA2)**, we can conclude that $G(t) = \mathbf{P}[C > t] \geq \mathbf{P}[C > L] > 0$ for all $t \in [0, L]$. Using this together with **(RA3)** we have

$$\begin{aligned} \mathbf{E}\left[\frac{\delta Z}{G(Z)}\middle| X\right] &= \mathbf{E}\left[I_{[Y<C]}\frac{Y}{G(Y)}\middle| X\right] = \mathbf{E}\left[\mathbf{E}\left[I_{[Y<C]} \,|\, X, Y\right]\frac{Y}{G(Y)}\middle| X\right] \\ &= \mathbf{E}[\,Y \mid X\,] \end{aligned}$$

and

$$\begin{aligned} \mathbf{E}\left[\int_0^Z \frac{dt}{G(t)}\middle| X\right] &= \mathbf{E}\left[\int_0^Y \frac{I_{[t<C]}}{G(t)}\,dt\middle| X\right] = \mathbf{E}\left[\int_0^Y \frac{\mathbf{E}\left[\,I_{[t<C]} \,\middle|\, X, Y\,\right]}{G(t)}\,dt\middle| X\right] \\ &= \mathbf{E}[\,Y \mid X\,]. \end{aligned}$$

Now, the last two equalities together with (2.9) imply (2.7).

Since in our case the survival function G of the censoring time is unknown, the random variables Y^*, $Y_1{}^*$, ..., $Y_n{}^*$ are not calculable. An obvious idea is to replace G in (2.9) and (2.10) by an estimate G_n, the well known Kaplan-Meier product-limit estimator (see, e.g., Kaplan and Meier (1958))

$$G_n(t) := \prod_{\substack{i=1,\ldots,n:\\ Z_{(i)}\leq t}} \left[\frac{n-i}{n-i+1}\right]^{1-\delta_{(i)}} \qquad (t \in \mathbb{R}) \tag{2.11}$$

$\left(0^0 := 1\right)$. Here, $(Z_{(i)}, \delta_{(i)})$, $i = 1, \ldots, n$, denote the observed pairs (Z_i, δ_i), arranged in such a way that

$$Z_{(1)} \leq Z_{(2)} \leq \ldots \leq Z_{(n)},$$

where in the case of ties censored observations ($\delta_i = 0$) occur before uncensored observations ($\delta_i = 1$):

$$Z_{(i)} = Z_{(j)}, \delta_{(i)} = 0, \delta_{(j)} = 1 \quad \Rightarrow \quad i < j.$$

This replacement results in

$$\hat{Y} := (1+\alpha) \int_0^Z \frac{dt}{G_n(t)} - \alpha \frac{\delta Z}{G_n(Z)}, \tag{2.12}$$

$$\hat{Y}_i := (1+\alpha) \int_0^{Z_i} \frac{dt}{G_n(t)} - \alpha \frac{\delta_i Z_i}{G_n(Z_i)} \quad (i = 1, \ldots, n) \tag{2.13}$$

(where we set $\frac{0}{0} := 0$), and

$$\hat{\mathcal{D}}_n := \left\{ (X_1, \hat{Y}_1), \ldots, (X_n, \hat{Y}_n) \right\} \tag{2.14}$$

Note that $\hat{Y}, \hat{Y}_1, \ldots, \hat{Y}_n$ depend on the sample size n and we have suppressed this in our notation. Furthermore, we want to stress that these random variables are in general neither independent nor identically distributed or even fulfill an equality similar to (2.7). The key step in the proof of our main results (vide Section 6) is rather to control the squared differences $|{Y_1}^* - \hat{Y}_1|^2, \ldots, |{Y_n}^* - \hat{Y}_n|^2$ (q.v. Section 4).

For the data $\hat{\mathcal{D}}_n$ we can now define multivariate smoothing spline estimates for censored regression analog to Definition 2.1. Let $d, k \in \mathbb{N}$ with $2k > d$ and $\lambda_n > 0$. Let $\hat{\mathcal{D}}_n$ be defined by (2.14). Our MSSE for censored data is given by

$$\tilde{m}_{n,(k,\lambda_n)}(\cdot) := \tilde{m}_{n,(k,\lambda_n)}(\cdot, \hat{\mathcal{D}}_n) := \underset{f \in W_k([0,1]^d)}{\arg\min} \left(\frac{1}{n} \sum_{i=1}^n |f(X_i) - \hat{Y}_i|^2 + \lambda_n J_k^2(f) \right) \tag{2.15}$$

with $W_k([0,1]^d)$ and $J_k^2(f)$ ($f \in W_k([0,1]^d)$) defined as in (2.3) and (2.5), respectively.

Since we assumed $0 \leq Y \leq L < \infty$ a.s., and therefore $0 \leq m(x) \leq L$ ($x \in [0,1]^d$), we truncate our estimate (2.15) such that it is bounded in the same way:

$$m_{n,(k,\lambda_n)}(\cdot) := T_{[0,L]} \tilde{m}_{n,(k,\lambda_n)}(\cdot), \tag{2.16}$$

where, for $a_1, a_2, t \in \mathbb{R}$ with $a_1 \leq a_2$,

$$T_{[a_1,a_2]} t := \begin{cases} a_2 & \text{if} \quad t > a_2 \\ t & \text{if} \quad a_1 \leq t \leq a_2 \\ a_1 & \text{if} \quad t < a_1, \end{cases}$$

and for all functions $f : \mathbb{R}^d \to \mathbb{R}$, we define $T_{[a_1,a_2]} f : \mathbb{R}^d \to \mathbb{R}$ by $\left(T_{[a_1,a_2]} f\right)(x_0) := T_{[a_1,a_2]}(f(x_0))$ $(x_0 \in \mathbb{R}^d)$.

2.3 Adaptation via splitting of the sample

The estimates (2.15) and (2.16) depend on the smoothing parameter λ_n and on k, which defines the degree of the Sobolev space $W_k([0,1]^d)$. Theorem 3.2 below gives a guideline for the choice of these parameters. But on the one hand, for λ_n, this is an asymptotic one which is orientated only towards the sample size n, not to the concrete realization of the sample. It is evident that a non-data-dependent choice of the smoothing parameter can lead to very unsatisfactory results. On the other hand, the choice of the parameters k and λ_n in Theorem 3.2 depend on the smoothness of the regression function which is unknown in a statistical application. Therefore we modify the estimate in a second step (c.v. Theorem 3.3) such that it adapts automatically to the smoothness of the regression function and choose k and λ_n in a totally data-dependent way via the splitting of the sample technique.

Let $n \geq 2$ and denote by $\lfloor t \rfloor$ and $\lceil t \rceil$ the integer part and the upper integer part of $t \in \mathbb{R}$, respectively. Consider the set of parameters $K_n \times \Lambda_n$ with

$$K_n := \left\{ \left\lfloor \frac{d}{2} \right\rfloor + 1, \left\lfloor \frac{d}{2} \right\rfloor + 2, \ldots, \left\lfloor \frac{d}{2} \right\rfloor + \left\lfloor (\ln n)^2 \right\rfloor \right\} \tag{2.17}$$

(where we define $K_2 := \left\{ \lfloor \frac{d}{2} \rfloor + 1 \right\}$ for $n = 2$) and

$$\Lambda_n := \left\{ \frac{\ln n}{2^n}, \frac{\ln n}{2^{n-1}}, \ldots, \frac{\ln n}{1} \right\}. \tag{2.18}$$

We split the sample (2.14) into two parts, the learning or training data

$$\hat{\mathcal{D}}_{n_1} := \left\{ (X_1, \hat{Y}_1), \ldots, (X_{n_1}, \hat{Y}_{n_1}) \right\}$$

and the testing data

$$\hat{\mathcal{D}}_{n_{\mathtt{t}}} := \left\{ (X_{n_1+1}, \hat{Y}_{n_1+1}), \ldots, (X_n, \hat{Y}_n) \right\},$$

with $n_{\mathtt{t}} + n_1 = n$.

For each pair of parameters $(k, \lambda) \in K_n \times \Lambda_n$ we first use the learning data to define an estimate $m_{n_1,(k,\lambda)}$ via

$$\tilde{m}_{n_1,(k,\lambda)}(\cdot) := \underset{f \in W_k([0,1]^d)}{\arg\min} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} |f(X_i) - \hat{Y}_i|^2 + \lambda J_k^2(f) \right),$$

where $W_k([0,1]^d)$ and $J_k^2(f)$ ($f \in W_k([0,1]^d)$ are given by (2.3) and (2.5), and

$$m_{n_1,(k,\lambda)}(\cdot) := T_{[0,L]} \tilde{m}_{n_1,(k,\lambda)}(\cdot). \tag{2.19}$$

Then we choose that estimate out of all calculated estimates (2.19) which performs best on the testing data in terms of the empirical $\mathcal{L}_2$ risk, i.e., our modified estimate is defined as

$$m_n(\cdot) := m_{n_1,(\hat{k},\hat{\lambda})}(\cdot), \tag{2.20}$$

where

$$\left(\hat{k},\hat{\lambda}\right) := \underset{(k,\lambda)\in K_n\times\Lambda_n}{\arg\min}\left(\frac{1}{n_{\mathrm{t}}}\sum_{i=n_1+1}^{n}|m_{n_1,(k,\lambda)}(X_i)-\hat{Y}_i|^2\right). \tag{2.21}$$

3 Main results

Now we are ready to present our main results, Theorem 3.1 – 3.3. Note that they are valid for any (fixed) $\alpha \in \mathbb{R}$, which is the parameter of the transformation of the censored data (vide Subsection 2.2).

Our first result states conditions on λ_n under which our multivariate smoothing spline estimates are strongly consistent for all distributions of (X, Y, C).

Theorem 3.1 (Consistency) *Let $k, d \in \mathbb{N}$ with $2k > d$ and $\alpha \in \mathbb{R}$. For $n \in \mathbb{N}$ choose $\lambda_n > 0$ such that $\lambda_n \to 0$ $(n \to \infty)$ and*

$$\frac{n\lambda_n^{\frac{d}{2k}}}{\ln n} \to \infty \quad (n \to \infty).$$

Let the estimate $m_{n,(k,\lambda_n)}$ be defined by (2.15) and (2.16). Then

$$\int_{\mathbb{R}^d} |m_{n,(k,\lambda_n)}(x) - m(x)|^2 \mu(dx) \to 0 \quad (n \to \infty) \quad a.s.$$

for every distribution of (X, Y, C) satisfying **(RA1)** – **(RA4)**.

Remark 3.1 Note that in Theorem 3.1 Assumption **(RA1)** can be abandoned if we slightly modify the estimate (vide Kohler and Krzyżak (2001), Remark 3).

Remark 3.2 It follows from Assumption **(RA2)**, the proof of Theorems 3.1– 3.3, and Corollary 1.3 in Stute and Wang (1993) that Assumption **(RA4)** can be dropped in Theorem 3.1 if we assume that F and G do not have common jumps and that either $\mathbf{P}[\,C = \tau_F\,] > 0$ or $\mathbf{P}[\,C = \tau_F\,] = 0$ but $\mathbf{P}[\,Y = \tau_F\,] > 0$.

In Theorems 3.2 – 3.3 we present our results concerning the rate of convergence of the MSSE for censored data. Therein, the following notation will be used: For two random variables $H_n, V_n \in \mathbb{R}_+$ we write $H_n = \mathcal{O}_{\mathcal{P}}(V_n)$, if there exists a constant $B > 0$ such that $\lim_{n\to\infty} \mathbf{P}[\,H_n > B \cdot V_n\,] = 0$. Along this line, the next theorem shows that our estimate achieves the optimal rate of convergence up to some logarithmic factor (q.v. Remark 3.3) for smooth regression functions $m \in W_p([0,1]^d)$, where $p \in \mathbb{N}$ with $2p > d$.

Theorem 3.2 (Rate of convergence) *Let $d, n \in \mathbb{N}$, $\alpha \in \mathbb{R}$, and $L \geq 1$. Let $p \in \mathbb{N}$ with $2p > d$ be arbitrary. If we choose the parameters k and λ_n of the estimate $m_{n,(k,\lambda_n)}$, which is defined by (2.15) and (2.16), such that $k = p$ and λ_n fulfills*

$$\lambda_n = b_1 \cdot \left(\frac{(\ln n)^2}{n}\right)^{\frac{2p}{2p+d}} J_p^2(m)^{-\frac{2p}{2p+d}}$$

with an arbitrary constant $b_1 > 0$, *then*

$$\int_{\mathbb{R}^d} |m_{n,(k,\lambda_n)}(x) - m(x)|^2 \mu(dx) = \mathcal{O}_{\mathcal{P}} \left(\left(\frac{(\ln n)^2}{n} \right)^{\frac{2p}{2p+d}} J_p^2(m)^{\frac{d}{2p+d}} \right)$$

for every distribution of (X, Y, C) *satisfying* **(RA1)** – **(RA4)**, $m \in W_p([0,1]^d)$ *with* $0 < J_p^2(m) < \infty$, *and*

$$-\int_0^{\tau_F} F(t)^{\frac{-p}{p+d}} dG(t) < \infty. \tag{3.1}$$

Note that since G is monotonically decreasing, the left hand side of (3.1)) is always non-negative.

Remark 3.3 Stone (1982) showed that the optimal rate of convergence (in adequate minimax sense) in nonparametric regression for estimates of (p, B)-smooth regression functions is given by $n^{-\frac{2p}{2p+d}}$. For $p \in \mathbb{N}$ and $B \in [0, \infty)$ a function $f : \mathbb{R}^d \to \mathbb{R}$ is called (p, B)-smooth if

$$\left| \frac{\partial^{(p-1)} f}{\partial x_1^{p_1} \ldots \partial x_d^{p_d}}(x) - \frac{\partial^{(p-1)} f}{\partial x_1^{p_1} \ldots \partial x_d^{p_d}}(x_0) \right| \leq B \cdot \|x - x_0\|$$

for all $p_1, \ldots, p_d \in \mathbb{N}_0$ with $p_1 + \ldots + p_d = p - 1$ and all $x, x_0 \in \mathbb{R}^d$. Since in our setting it is allowed that no censoring arises (i.e., $\mathbf{P}[C > L] = 1$, vide Assumption **(RA2)**), we deduce that the rate of convergence in Theorem 3.2 is optimal up to the logarithmic factor $(\ln n)^{\frac{4p}{2p+d}}$. Note that the rate in Theorem 3.2 is identical to known rates for MSSE in nonparametric regression with random design (q.v. Kohler, Krzyżak, and Schäfer (2002)). However, for censored regression, the additional assumptions on the distribution of C are needed.

Assume that $m \in W_p([0,1]^d)$ for some $p \in \mathbb{N}$ with $2p > d$. In Theorem 3.2, to achieve the nearly optimal rate of convergence, the parameters k and λ_n of our estimate (2.16) have to be chosen such that they depend on the smoothness of the regression function m, measured by p and $J_p^2(m)$. Since in practical applications the smoothness of m is unknown, these parameters and hence the estimate cannot be calculated . Therefore it is necessary to apply adaptation procedures which allow a completely data-driven choice of the parameters without loosing the properties of Theorem 3.2. The next theorem uses the splitting of the sample technique (vide Section 2.3).

Theorem 3.3 (Adaptation via splitting of the sample) *Let* $d, n \in \mathbb{N}$ *with* $n \geq 2$ *and set* $n_1 := \lceil \frac{n}{2} \rceil$. *Let the set of parameters* $K_n \times \Lambda_n$ *be defined by (2.17) and (2.18). Let* $L \geq 1$, $\alpha \in \mathbb{R}$, *and the estimate* m_n *be given by (2.20). For any* $p \in \mathbb{N}$ *with* $2p > d$, *we have*

$$\int_{\mathbb{R}^d} |m_n(x) - m(x)|^2 \mu(dx) = \mathcal{O}_{\mathcal{P}} \left(\left(\frac{(\ln n)^2}{n} \right)^{\frac{2p}{2p+d}} \left(J_p^2(m) \right)^{\frac{d}{2p+d}} \right)$$

for every distribution of (X,Y,C) *satisfying* **(RA1)** – **(RA4)***,* $m \in W_p([0,1]^d)$ *with* $0 < J_p^2(m) < \infty$*, and (3.1).*

Remark 3.4 The definition of the estimate m_n does not depend on p or $J_p^2(m)$, hence it automatically adapts to the unknown smoothness of the regression function. Besides, the rate of convergence is identical to that of Theorem 3.2.

Remark 3.5 We want to stress that in Theorems 3.1 – 3.3 no assumption on the underlying distribution of X besides **(RA1)** is required. Especially, it is not required that X has a density with respect to the Lebesgue–Borel measure.

Remark 3.6 It follows from the proof of Theorems 3.1 – 3.3 and the Remark in Chen and Lo (1997), that Assumption **(RA4)** can be abandoned in Theorem 3.2 and Theorem 3.3. In this case, replace G and F in (3.1) by continuous survival functions $\tilde{G}$ and $\tilde{F}$, where $\tilde{G}$ smoothes the probability mass of G at its discontinuity points to small intervals and $\tilde{F}$ assigns probability 0 to these intervals. For details, see Chen and Lo (1997). Furthermore, one can conclude that if there exists an $q \in (0,1)$ (e.g., $q = \frac{2p}{2p+d}$ with $p, d \in \mathbb{N}$) such that

$$- \int_0^{\tau_F} \tilde{F}(t)^{\frac{-q}{2-q}} d\tilde{G}(t) < \infty$$

then Theorem 3.1 also holds if Assumption **(RA4)** is violated.

Remark 3.7 It follows from Corollary 2.2 in Chen and Lo (1997) and **(RA2)** that assumption (3.1) in Theorems 3.2 – 3.3 is fulfilled if there exists some $\beta \in \left(0, 1 + \frac{d}{p}\right)$ such that

$$\limsup_{t \to \tau_F} \frac{(G(t) - G(\tau_F))^\beta}{F(t)} < \infty.$$

The proofs of Theorems 3.1–3.3 are given in Section 6.

4 MSSE applied to data with additional measurement errors in the dependent variable

Here we shall put the setting of Subsection 2.2 in a more general context. Let therefore be $(X, Y^\star) \in [0,1]^d \times [-L^\star, L^\star]$ a.s. a random vector where $L^\star \in [0, \infty)$. In some situations, data from the distribution of $(X, Y^\star)$ can only be observed with additional measurement errors in the dependent variable (see, e.g., Kohler (2002)). Here, we do not assume that these errors are independent or identically distributed.

So in order to calculate an estimate of the regression function $m(x) = \mathbf{E}\left[Y^\star \,|\, X = x \right]$, one has only given the data

$$\bar{\mathcal{D}}_n := \left\{ (X_1, \bar{Y}_{1,n}), \ldots, (X_n, \bar{Y}_{n,n}) \right\}, \tag{4.1}$$

instead of a sample $(X_1, Y_1{}^\star), \dots, (X_n, Y_n{}^\star)$ of i.i.d. copies of $(X, Y^\star)$. In the following, we shall suppress the dependency of $\bar{Y}_{1,n}, \dots, \bar{Y}_{n,n}$ on the sample size n, i.e., use the notation $\bar{Y}_i = \bar{Y}_{i,n}$.

Note that $Y^*, Y_1{}^*, \dots, Y_n{}^*$ and $\hat{Y}_1, \dots, \hat{Y}_n$ as defined in Section 2.2 are special choices of $Y^\star, Y_1{}^\star, \dots, Y_n{}^\star$ and $\bar{Y}_1, \dots, \bar{Y}_n$, respectively. Nevertheless, for the sake of generality, throughout this section we do not demand that either $Y^\star, Y_1{}^\star, \dots, Y_n{}^\star$ or $\bar{Y}_1, \dots, \bar{Y}_n$ take a special form, or are even the result of a censoring unbiased transformation. Instead, the only assumption needed here beside $Y^\star \in [-L^\star, L^\star]$ a.s, is that the squared measurement errors $|Y_1{}^\star - \bar{Y}_1|^2, \dots, |Y_n{}^\star - \bar{Y}_n|^2$ are "small".

The MSSE considered in this section differ from those defined in Subsections 2.2 and 2.3 only in that way that we use the data (4.1) instead of (2.14) and a truncation at $[-L^\star, L^\star]$ instead of $[0, L]$ in order to define them. To be more precise, the estimates are now given by

$$\tilde{m}_{n,(k,\lambda_n)}(\cdot) := \tilde{m}_{n,(k,\lambda_n)}(\cdot, \bar{\mathcal{D}}_n) := \underset{f \in W_k([0,1]^d)}{\arg\min} \left(\frac{1}{n} \sum_{i=1}^{n} |f(X_i) - \bar{Y}_i|^2 + \lambda_n J_k^2(f) \right), \quad (4.2)$$

with $2k > d$, $\lambda_n > 0$, $W_k([0,1]^d)$ and $J_k^2(\cdot)$ defined as in Definition 2.1, and

$$m_{n,(k,\lambda_n)}(\cdot) := T_{[-L^\star, L^\star]} \tilde{m}_{n,(k,\lambda_n)}(\cdot). \quad (4.3)$$

Furthermore, our adaptive MSSE is defined by

$$m_n(\cdot) := m_{n_1,(\bar{k},\bar{\lambda})}(\cdot), \quad (4.4)$$

where

$$(\bar{k}, \bar{\lambda}) := \underset{(k,\lambda) \in K_n \times \Lambda_n}{\arg\min} \left(\frac{1}{n_{\mathtt{t}}} \sum_{i=n_1+1}^{n} |m_{n_1,(k,\lambda)}(X_i) - \bar{Y}_i|^2 \right).$$

Here $n_1, n_{\mathtt{t}}, K_n$ and Λ_n are defined as in Subsection 2.3.

In this setting the following three results hold.

Theorem 4.1 (Consistency) *Let $k, d \in \mathbb{N}$ with $2k > d$. For $n \in \mathbb{N}$ choose $\lambda_n > 0$ such that $\lambda_n \to 0$ $(n \to \infty)$ and*

$$\frac{n \lambda_n^{\frac{d}{2k}}}{\ln n} \to \infty \quad (n \to \infty).$$

Let the estimate $m_{n,(k,\lambda_n)}$ be defined by (4.2) and (4.3). If

$$\frac{1}{n} \sum_{i=1}^{n} |Y_i^\star - \bar{Y}_i|^2 \to 0 \quad (n \to \infty) \quad a.s., \quad (4.5)$$

then

$$\int_{\mathbb{R}^d} |m_{n,(k,\lambda_n)}(x) - m(x)|^2 \mu(dx) \to 0 \quad (n \to \infty) \quad a.s. \quad (4.6)$$

for every distribution of $(X, Y^\star)$ with $(X, Y^\star) \in [0,1]^d \times [-L^\star, L^\star]$ a.s.

Theorem 4.2 (Rate of convergence) *Let $d, n \in \mathbb{N}$, and $L^\star \geq 1$. Let $p \in \mathbb{N}$ with $2p > d$ be arbitrary. Assume that we have chosen the parameters k and λ_n of the estimate $m_{n,(k,\lambda_n)}$ which is defined by (4.2) and (4.3), such that $k = p$ and λ_n fulfills*

$$\lambda_n = b_1 \cdot \left(\frac{(\ln n)^2}{n} \right)^{\frac{2p}{2p+d}} J_p^2(m)^{-\frac{2p}{2p+d}} \tag{4.7}$$

with an arbitrary constant $b_1 > 0$. If there exists a constant $b_2 > 0$ such that

$$\mathbf{P} \left[\max_{i=1,\ldots,n} |Y_i^\star - \bar{Y}_i|^2 > b_2 \right] \to 0 \quad (n \to \infty), \tag{4.8}$$

then we have

$$\int_{\mathbb{R}^d} |m_{n,(k,\lambda_n)}(x) - m(x)|^2 \mu(dx) = \mathcal{O}_{\mathcal{P}} \left(\frac{1}{n} \sum_{i=1}^n |Y_i^\star - \bar{Y}_i|^2 + \left(\frac{(\ln n)^2}{n} \right)^{\frac{2p}{2p+d}} J_p^2(m)^{\frac{d}{2p+d}} \right)$$

for every distribution of $(X, Y^\star)$ with $(X, Y^\star) \in [0,1]^d \times [-L^\star, L^\star]$ a.s., $m \in W_p([0,1]^d)$, and $0 < J_p^2(m) < \infty$.

Theorem 4.3 (Adaptation via splitting of the sample) *Let $d, n \in \mathbb{N}$ with $n \geq 2$ and set $n_1 := \lceil \frac{n}{2} \rceil$. Let the set of parameters $K_n \times \Lambda_n$ be defined by (2.17) and (2.18). Let $L^\star \geq 1$ and the estimate m_n be given by (4.4). Assume that there exists a constant $b_2 > 0$ such that (4.8) holds. For any $p \in \mathbb{N}$ with $2p > d$, we have*

$$\int_{\mathbb{R}^d} |m_n(x) - m(x)|^2 \mu(dx) = \mathcal{O}_{\mathcal{P}} \left(\frac{1}{n} \sum_{i=1}^n |Y_i^\star - \bar{Y}_i|^2 + \left(\frac{(\ln n)^2}{n} \right)^{\frac{2p}{2p+d}} \left(J_p^2(m) \right)^{\frac{d}{2p+d}} \right)$$

for every distribution of $(X, Y^\star)$ with $(X, Y^\star) \in [0,1]^d \times [-L^\star, L^\star]$ a.s., $m \in W_p([0,1]^d)$, and $0 < J_p^2(m) < \infty$.

We shall use Theorems 4.1 – 4.3 in order to prove Theorems 3.1 – 3.3 in Section 6.

5 Proof of Theorems 4.1–4.3

In this section, our results from nonparametric regression analysis with additional measurement errors in the dependent variable are proven. In the proofs of Theorems 4.1 and 4.2, we need the concept of covering numbers.

Definition 5.1 (Covering number) *Let $d \in \mathbb{N}$, $1 \leq r < \infty$, and $\mathcal{F}$ be a class of functions $f : \mathbb{R}^d \to \mathbb{R}$. For any $\epsilon > 0$ and any $v_1^n = (v_1, \ldots, v_n) \in (\mathbb{R}^d)^n$, the* **covering number** *$\mathcal{N}_r(\epsilon, \mathcal{F}, v_1^n)$ is defined as the smallest integer N such that there exist functions $g_1, \ldots, g_N : \mathbb{R}^d \to \mathbb{R}$ with*

$$\min_{1 \leq j \leq N} \left(\frac{1}{n} \sum_{i=1}^n |f(v_i) - g_j(v_i)|^r \right)^{\frac{1}{r}} \leq \epsilon$$

for each $f \in \mathcal{F}$.

PROOF OF THEOREM 4.1. First, we shall prove the following lemma

Lemma 5.1 *Let $k, d \in \mathbb{N}$ with $2k > d$ and $L^\star \in \mathbb{R}_+$. For $n \in \mathbb{N}$ choose $\lambda_n > 0$ such that $\lambda_n \to 0$ $(n \to \infty)$ and*

$$\frac{n\lambda_n^{\frac{d}{2k}}}{\ln n} \to \infty \quad (n \to \infty). \tag{5.1}$$

Let the estimate $m_{n,(k,\lambda_n)}$ be defined by (4.2) and (4.3). If

$$\frac{1}{n}\sum_{i=1}^{n} |Y_i^\star - \bar{Y}_i|^2 \to 0 \quad (n \to \infty) \quad a.s. \tag{5.2}$$

then

$$\mathbf{E}\left[\,\left|m_{n,(k,\lambda_n)}(X) - Y^\star\right|^2 \middle| \bar{\mathcal{D}}_n\right] - \frac{1}{n}\sum_{i=1}^{n}\left|m_{n,(k,\lambda_n)}(X_i) - Y_i^\star\right|^2 \to 0 \quad (n \to \infty) \quad a.s. \tag{5.3}$$

for every distribution of $(X, Y^\star)$ with $(X, Y^\star) \in [0,1]^d \times [-L^\star, L^\star]$ a.s.

PROOF OF LEMMA 5.1. By the strong law of large numbers, Definition (4.2) of the estimate $\tilde{m}_{n,(k,\lambda_n)}$, and (5.2)

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left|\tilde{m}_{n,(k,\lambda_n)}(X_i) - \bar{Y}_i\right|^2 + \lambda_n J_k^2(\tilde{m}_{n,(k,\lambda_n)}) \le \frac{1}{n}\sum_{i=1}^{n}\left|\bar{Y}_i\right|^2 \\
&\le \frac{2}{n}\sum_{i=1}^{n}\left|Y_i^\star - \bar{Y}_i\right|^2 + \frac{2}{n}\sum_{i=1}^{n}|Y_i^\star|^2 \to 2\mathbf{E}(Y^\star)^2 \quad (n \to \infty) \quad \text{a.s.}
\end{aligned}$$

This implies that with probability 1, for sufficiently large n,

$$m_{n,(k,\lambda_n)} \in \mathcal{F}_{3(L^\star)^2/\lambda_n} := \left\{T_{[-L^\star, L^\star]}f : f \in W_k([0,1]^d), J_k^2(f) \le \frac{3(L^\star)^2}{\lambda_n}\right\}$$

Thus it suffices to show

$$\sup_{g \in \mathcal{G}_{3(L^\star)^2/\lambda_n}} \left|\mathbf{E}\left[g(X, Y^\star)\right] - \frac{1}{n}\sum_{i=1}^{n} g(X_i, Y_i^\star)\right| \to 0 \quad (n \to \infty) \quad \text{a.s.}$$

where

$$\mathcal{G}_{3(L^\star)^2/\lambda_n} := \left\{g : g(x,y) = |f(x) - y|^2, f \in \mathcal{F}_{3(L^\star)^2/\lambda_n}, x \in \mathbb{R}^d, y \in [-L^\star, L^\star]\right\}.$$

For this purpose, we first note that for two functions $g_1, g_2 \in \mathcal{G}_{3(L^\star)^2/\lambda_n}$ with $g_j(x,y) = |f_j(x) - y|^2$, $f_j \in \mathcal{F}_{3(L^\star)^2/\lambda_n}$, $j = 1, 2$, we have

$$\frac{1}{n}\sum_{i=1}^{n}|g_1(X_i, Y_i^\star) - g_2(X_i, Y_i^\star)| \le 4L^\star \frac{1}{n}\sum_{i=1}^{n}|f_1(X_i) - f_2(X_i)| \quad \text{a.s.}$$

which implies for all $\epsilon > 0$

$$\mathcal{N}_1\left(\frac{\epsilon}{8}, \mathcal{G}_{3(L^\star)^2/\lambda_n}, (X, Y^\star)_1^n\right) \leq \mathcal{N}_1\left(\frac{\epsilon}{32L^\star}, \mathcal{F}_{3(L^\star)^2/\lambda_n}, X_1^n\right) \quad \text{a.s.}$$

The last inequality, Theorem 9.1 in Györfi, Kohler, Krzyżak, and Walk (2002), Lemma 3 in Kohler and Krzyżak (2001), and (5.1) imply that for all $0 < \epsilon < 32(L^\star)^2$ and all sufficiently large n

$$\begin{aligned}
&\mathbf{P}\left[\sup_{g\in\mathcal{G}_{3(L^\star)^2/\lambda_n}}\left|\mathbf{E}\left[g(X,Y^\star)\right] - \frac{1}{n}\sum_{i=1}^n g(X_i, Y_i^\star)\right| > \epsilon\right] \\
&\leq\ 8\exp\left(-\frac{n\epsilon^2}{128\left(4(L^\star)^2\right)^2}\right)\cdot \mathbf{E}\,\mathcal{N}_1\left(\frac{\epsilon}{32L^\star}, \mathcal{F}_{3(L^\star)^2/\lambda_n}, X_1^n\right) \\
&\leq\ 8\exp\left(-\frac{n\epsilon^2}{2048(L^\star)^4} + \left[B_1\left(\sqrt{\frac{3(L^\star)^2}{\lambda_n}}\frac{32L^\star}{\epsilon}\right)^{\frac{d}{k}} + B_2\right]\ln\left(B_3\frac{32(L^\star)^2 n}{\epsilon}\right)\right) \\
&\leq\ 8\exp\left(-\frac{1}{2}\frac{n\epsilon^2}{2048(L^\star)^4}\right),
\end{aligned}$$

where $B_1, B_2, B_3 > 0$ are constants which only depend on k and d. From this, the assertion of Lemma 5.1 follows by an application of the Borel–Cantelli lemma. □

Now we start with the proof of Theorem 4.1. Let $\epsilon > 0$ be arbitrary. From Theorem 3.14 in Rudin (1974), one can conclude that there exists a function $g_\epsilon \in W_k([0,1]^d)$ such that

$$\int_{\mathbb{R}^d} |g_\epsilon(x) - m(x)|^2\,\mu(dx) \leq \epsilon \quad \text{and} \quad J_k^2(g_\epsilon) < \infty. \tag{5.4}$$

In this sequel, the following error decomposition will be used

$$\begin{aligned}
&\int_{\mathbb{R}^d}\left|m_{n,(k,\lambda_n)}(x) - m(x)\right|^2\mu(dx) \\
&=\ \mathbf{E}\left[\left|m_{n,(k,\lambda_n)}(X) - Y^\star\right|^2\middle|\bar{\mathcal{D}}_n\right] - \mathbf{E}\left[\left|m(X) - Y^\star\right|^2\right] \\
&=:\ \sum_{j=1}^8 H_{j,n}.
\end{aligned}$$

Below we show how to bound each of these eight terms from above . By an application of Lemma 5.1, we get

$$H_{1,n} := \mathbf{E}\left[\left|m_{n,(k,\lambda_n)}(X) - Y^\star\right|^2\middle|\bar{\mathcal{D}}_n\right] - \frac{1}{n}\sum_{i=1}^n\left|m_{n,(k,\lambda_n)}(X_i) - Y_i^\star\right|^2 \to 0 \quad (n\to\infty) \quad \text{a.s.}$$

The definition of the truncated estimate (4.3) and $|Y^\star| \leq L^\star$ imply

$$H_{2,n} := \frac{1}{n}\sum_{i=1}^n\left|m_{n,(k,\lambda_n)}(X_i) - Y_i^\star\right|^2 - \frac{1}{n}\sum_{i=1}^n\left|\tilde{m}_{n,(k,\lambda_n)}(X_i) - Y_i^\star\right|^2 \leq 0$$

In order to bound the third term

$$H_{3,n} := \frac{1}{n}\sum_{i=1}^{n}\left|\tilde{m}_{n,(k,\lambda_n)}(X_i) - Y_i^\star\right|^2 - (1+\epsilon)\frac{1}{n}\sum_{i=1}^{n}\left|\tilde{m}_{n,(k,\lambda_n)}(X_i) - \bar{Y}_i\right|^2$$

from above, observe that for all $a, b \geq 0$, we have

$$(a+b)^2 \leq a^2(1+\epsilon) + b^2\left(1+\frac{1}{\epsilon}\right). \tag{5.5}$$

From (4.5) and (5.5) one can conclude

$$H_{3,n} \leq \left(1+\frac{1}{\epsilon}\right)\frac{1}{n}\sum_{i=1}^{n}\left|Y_i^\star - \bar{Y}_i\right|^2 \to 0 \quad (n\to\infty) \quad \text{a.s.}$$

From (5.4), the definition of $\tilde{m}_{n,(k,\lambda_n)}$, and $\lambda_n \to 0$ $(n\to\infty)$ follows

$$\begin{aligned} H_{4,n} &:= (1+\epsilon)\left[\frac{1}{n}\sum_{i=1}^{n}\left|\tilde{m}_{n,(k,\lambda_n)}(X_i) - \bar{Y}_i\right|^2 - \frac{1}{n}\sum_{i=1}^{n}\left|g_\epsilon(X_i) - \bar{Y}_i\right|^2\right] \\ &\leq (1+\epsilon)\,\lambda_n J_k^2(g_\epsilon) \to 0 \quad (n\to\infty) \end{aligned}$$

Using again (4.5) and (5.5), we have

$$\begin{aligned} H_{5,n} &:= (1+\epsilon)\left[\frac{1}{n}\sum_{i=1}^{n}\left|g_\epsilon(X_i) - \bar{Y}_i\right|^2 - (1+\epsilon)\frac{1}{n}\sum_{i=1}^{n}\left|g_\epsilon(X_i) - Y_i^\star\right|^2\right] \\ &\leq (1+\epsilon)\left(1+\frac{1}{\epsilon}\right)\frac{1}{n}\sum_{i=1}^{n}\left|Y_i^\star - \bar{Y}_i\right|^2 \to 0 \quad (n\to\infty) \quad \text{a.s.} \end{aligned}$$

and by the strong law of large numbers

$$H_{6,n} := (1+\epsilon)^2\left[\frac{1}{n}\sum_{i=1}^{n}|g_\epsilon(X_i) - Y_i^\star|^2 - \mathbf{E}\left[|g_\epsilon(X) - Y^\star|^2\right]\right] \to 0 \quad (n\to\infty) \quad \text{a.s.}$$

An application of (5.4) yields

$$\begin{aligned} H_{7,n} &:= (1+\epsilon)^2\,\mathbf{E}\left[|g_\epsilon(X) - Y^\star|^2\right] - (1+\epsilon)^2\,\mathbf{E}\left[|m(X) - Y^\star|^2\right] \\ &\leq \epsilon\,(1+\epsilon)^2. \end{aligned}$$

Finally, we get the following upper bound for the last of the eight terms

$$H_{8,n} := ((1+\epsilon)^2 - 1)\,\mathbf{E}\left[|m(X) - Y^\star|^2\right] \leq ((1+\epsilon)^2 - 1)\,(2L^\star)^2.$$

Combining all the results from above, one can conclude

$$\limsup_{n\to\infty}\int_{\mathbb{R}^d}|m_{n,(k,\lambda_n)}(x) - m(x)|^2\mu(dx) \leq \epsilon\,(1+\epsilon)^2 + 4\,((1+\epsilon)^2 - 1)\,(L^\star)^2 \quad \text{a.s.}$$

With $\epsilon \to 0$ the assertion follows. $\square$

Proof of Theorem 4.2. Assume $k = p$. Let $B_1 > 0$ be an arbitrary constant and $b_2 > 0$ a sufficiently large constant. We show that for all $t \geq B_1 \lambda_n J_p^2(m)$ with $\lambda_n > 0$, $\lambda_n \to 0$ $(n \to \infty)$, and

$$\left(\frac{\ln n}{n}\right)^{\frac{2p}{2p+d}} \lambda_n^{-1} \to 0 \quad (n \to \infty), \tag{5.6}$$

we have

$$\begin{aligned} &\mathbf{P}\left[\int_{\mathbb{R}^d} \left|m_{n,(k,\lambda_n)}(x) - m(x)\right|^2 \mu(dx) > 3\,t + 4\lambda_n J_p^2(m) + \frac{128}{n}\sum_{i=1}^n \left|Y_i^\star - \bar{Y}_i\right|^2\right] \\ &\quad \leq 50 \frac{\exp\left(-B_2\, n\, t\right)}{1 - \exp\left(-B_2\, n\, t\right)} + b_5 \exp(-b_6\, n\, t) + \mathbf{P}\left[\max_{i=1,\dots,n} \left|Y_i^\star - \bar{Y}_i\right|^2 > b_2\right], \end{aligned} \tag{5.7}$$

where $B_2, b_5, b_6 > 0$ are constants which only depend on $L^\star$. Since (4.7) implies (5.6), $\lambda_n \to 0$ $(n \to \infty)$, and

$$n\,\lambda_n \to \infty \quad (n \to \infty), \tag{5.8}$$

the assertion of Theorem 4.2 follows from this together with (4.8) and (5.7).

In order to show (5.7), we first note that for all $t > 0$

$$\begin{aligned} &\mathbf{P}\left[\int_{\mathbb{R}^d} \left|m_{n,(k,\lambda_n)}(x) - m(x)\right|^2 \mu(dx) > 3t + 4\lambda_n J_p^2(m) + \frac{128}{n}\sum_{i=1}^n \left|Y_i^\star - \bar{Y}_i\right|^2\right] \\ &\quad \leq \mathbf{P}\left[H_{1,n} > t\right] + \mathbf{P}\left[H_{2,n} > t\right], \end{aligned} \tag{5.9}$$

where

$$\begin{aligned} H_{1,n} \;&:=\; \int_{\mathbb{R}^d} \left|m_{n,(k,\lambda_n)}(x) - m(x)\right|^2 \mu(dx) \\ &\quad -2\left[\frac{1}{n}\sum_{i=1}^n \left|m_{n,(k,\lambda_n)}(X_i) - m(X_i)\right|^2 + \lambda_n J_k^2(\tilde{m}_{n,(k,\lambda_n)})\right] \end{aligned}$$

and

$$\begin{aligned} H_{2,n} \;&:=\; \frac{1}{n}\sum_{i=1}^n \left|m_{n,(k,\lambda_n)}(X_i) - m(X_i)\right|^2 + \lambda_n J_k^2(\tilde{m}_{n,(k,\lambda_n)}) \\ &\quad -\frac{64}{n}\sum_{i=1}^n \left|Y_i^\star - \bar{Y}_i\right|^2 - 2\lambda_n J_p^2(m). \end{aligned}$$

In the following, upper bounds for each of the probabilities on the right hand side of (5.9) will be computed. By an application of the Peeling-technique (cf. Section 5.3 in van de Geer (2000)), we have for all $t > 0$

$$\mathbf{P}\left[H_{1,n} > t\right] \leq \sum_{j=0}^\infty \mathbf{P}\left[2^j t \leq 2\lambda_n J_k^2(\tilde{m}_{n,(k,\lambda_n)}) + t < 2^{j+1} t, H_{1,n} > t\right]. \tag{5.10}$$

For every $j = 0, 1, \ldots$ and all $t > 0$ set

$$H_{3,n,j}(t) := \frac{\mathbf{E}\left[\left|m_{n,(k,\lambda_n)}(X) - m(X)\right|^2 \middle| \bar{\mathcal{D}}_n\right] - \frac{1}{n}\sum_{i=1}^n \left|m_{n,(k,\lambda_n)}(X_i) - m(X_i)\right|^2}{\mathbf{E}\left[\left|m_{n,(k,\lambda_n)}(X) - m(X)\right|^2 \middle| \bar{\mathcal{D}}_n\right] + 2^j t}.$$

Then we can conclude from (5.10) for all $t > 0$

$$\begin{aligned}\mathbf{P}\left[H_{1,n} > t\right] &\leq \sum_{j=0}^{\infty} \mathbf{P}\left[J_k^2(\tilde{m}_{n,(k,\lambda_n)}) < \frac{2^j t}{\lambda_n}, H_{3,n,j}(t) > \frac{1}{2}\right] \\ &\leq \sum_{j=0}^{\infty} \mathbf{P}\left[\sup_{g \in \mathcal{G}_{2^j t/\lambda_n}} \frac{\mathbf{E}\, g(X) - \frac{1}{n}\sum_{i=1}^n g(X_i)}{\mathbf{E}\, g(X) + 2^j t} > \frac{1}{2}\right] \qquad (5.11)\end{aligned}$$

where for every $j = 0, 1, \ldots$

$$\mathcal{G}_{2^j t/\lambda_n} := \left\{g : g(x) = |f(x) - m(x)|^2, f \in \mathcal{F}_{2^j t/\lambda_n}, x \in [0,1]^d\right\}$$

with

$$\mathcal{F}_{2^j t/\lambda_n} := \left\{T_{[-L^\star, L^\star]} f : f \in W_k([0,1]^d), J_k^2(f) \leq \frac{2^j t}{\lambda_n}\right\}.$$

Fix $j = 0, 1, \ldots$. First, we note that for all $g_1, g_2 \in \mathcal{G}_{2^j t/\lambda_n}$ with $g_i(x) = |f_i(x) - m(x)|^2$, $f_i \in \mathcal{F}_{2^j t/\lambda_n}$, $i = 1, 2$ ($x \in [0,1]^d$) and all $x_1, \ldots, x_n \in [0,1]^d$, we have

$$\frac{1}{n}\sum_{i=1}^n |g_1(x_i) - g_2(x_i)|^2 \leq (4L^\star)^2 \frac{1}{n}\sum_{i=1}^n |f_1(x_i) - f_2(x_i)|^2$$

and therefore for all $s > 0$ and all $x_1, \ldots, x_n \in [0,1]^d$

$$\mathcal{N}_2\left(s, \mathcal{G}_{2^j t/\lambda_n}, x_1^n\right) \leq \mathcal{N}_2\left(\frac{s}{4L^\star}, \mathcal{F}_{2^j t/\lambda_n}, x_1^n\right) \qquad (5.12)$$

From (5.8) one can conclude for all $t \geq B_1 \lambda_n J_p^2(m)$, all $\zeta \geq \frac{2^j t}{4}$, and all sufficiently large n that $n\zeta \geq (L^\star)^2$. This together with Lemma B.2 and (5.12) yields

$$\begin{aligned}\int_0^{\sqrt{\zeta}} \sqrt{\ln \mathcal{N}_2\left(s, \mathcal{G}_{2^j t/\lambda_n}, x_1^n\right)}\, ds &\leq b_8\,(8L^\star)^{\frac{d}{2p}}\,\lambda_n^{-\frac{d}{4p}}\sqrt{\zeta}\sqrt{\ln n} + b_9\sqrt{\zeta}\sqrt{\ln n} \\ &\leq \frac{2\sqrt{n}\,\zeta}{\sqrt{B_1 J_p^2(m)}}\left(b_8\,(8L^\star)^{\frac{d}{2p}}\sqrt{\frac{\ln n}{n}\lambda_n^{-\frac{2p+d}{2p}}} + b_9\sqrt{\frac{\ln n}{n\lambda_n}}\right) \\ &\leq \frac{2\left(b_8\,(8L^\star)^{\frac{d}{2p}} + b_9\right)\sqrt{n}\,\zeta}{\sqrt{B_1 J_p^2(m)}}\sqrt{\left(\frac{\ln n}{n}\right)^{\frac{2p}{2p+d}}\lambda_n^{-1}} \qquad (5.13)\end{aligned}$$

for all $t \geq B_1 \lambda_n J_p^2(m)$, all $\zeta \geq \frac{2^j t}{4}$, all $x_1, \ldots, x_n \in [0,1]^d$, and all sufficiently large n with some constants $b_8, b_9 > 0$ which only depend on p and d. The last inequality, (5.6), (5.11), and Theorem 2 in Kohler (2000) (set there $\mathcal{X} := [0,1]^d$, $\mathcal{F} := \mathcal{G}_{2^j t/\lambda_n}$, $K_1 := K_2 := 4(L^\star)^2$, $\epsilon := \frac{1}{2}$, and $\alpha := 2^j t$) imply for all $t \geq B_1 \lambda_n J_p^2(m)$ and all sufficiently large n

$$\mathbf{P}\left[H_{1,n} > t \right] \leq 50 \sum_{j=0}^{\infty} \exp\left(-2^j B_2\, n\, t\right) \leq 50 \frac{\exp\left(-B_2\, n\, t\right)}{1 - \exp\left(-B_2\, n\, t\right)} \tag{5.14}$$

with a constant B_2, which only depends on $L^\star$.

Set $l := L^\star + \sqrt{b_2}$ and

$$m^\star_{n,(k,\lambda_n)}(\cdot) := T_{[-l,l]} \tilde{m}_{n,(k,\lambda_n)}(\cdot). \tag{5.15}$$

Then one can conclude for all $t > 0$

$$\mathbf{P}\left[H_{2,n} > t \right] \leq \mathbf{P}\left[H_{2,n} > t, \max_{i=1,\ldots,n} \left|\bar{Y}_i\right| \leq l \right] + \mathbf{P}\left[\max_{i=1,\ldots,n} \left|\bar{Y}_i\right| > l \right] =: q_{1,n} + q_{2,n}. \tag{5.16}$$

For the second term on the right hand side of (5.16), we have

$$q_{2,n} \leq \mathbf{P}\left[\max_{i=1,\ldots,n} \left|Y_i^\star - \bar{Y}_i\right| + L^\star > l \right] = \mathbf{P}\left[\max_{i=1,\ldots,n} \left|Y_i^\star - \bar{Y}_i\right|^2 > b_2 \right] \tag{5.17}$$

and for the first term (4.2), (4.3), and (5.15) imply for all $t > 0$

$$\begin{aligned} q_{1,n} \;\; &\leq \;\; \mathbf{P}\left[H_{2,n} > t, \frac{1}{n}\sum_{i=1}^n \left| m^\star_{n,(k,\lambda_n)}(X_i) - \bar{Y}_i \right|^2 \leq \frac{1}{n}\sum_{i=1}^n \left| \tilde{m}_{n,(k,\lambda_n)}(X_i) - \bar{Y}_i \right|^2 \right] \\ &\leq \;\; \mathbf{P}\Bigg[\frac{1}{n}\sum_{i=1}^n \left| m^\star_{n,(k,\lambda_n)}(X_i) - m(X_i) \right|^2 + \lambda_n J_k^2(\tilde{m}_{n,(k,\lambda_n)}) \\ &\qquad\qquad > t + \frac{64}{n}\sum_{i=1}^n \left|Y_i^\star - \bar{Y}_i\right|^2 + 2\lambda_n J_p^2(m), \\ &\qquad \frac{1}{n}\sum_{i=1}^n \left| m^\star_{n,(k,\lambda_n)}(X_i) - \bar{Y}_i \right|^2 \leq \frac{1}{n}\sum_{i=1}^n \left| \tilde{m}_{n,(k,\lambda_n)}(X_i) - \bar{Y}_i \right|^2 \Bigg], \end{aligned} \tag{5.18}$$

where the last inequality follows from $|m(X)| \leq L^\star < l$ a.s. Next, we shall apply Lemma A.1 (note that we have chosen k such that $k = p$). Since $\lambda_n \to 0$ ($n \to \infty$) and 5.6 imply (A.4) – (A.6), one can conclude from Lemma A.1 for all sufficiently large n and all $t \geq B_1 \lambda_n J_p^2(m)$

$$q_{1,n} \leq b_5 \exp(-b_6 n t)$$

with some constants $b_5, b_6 > 0$ which only depend on $L^\star$. This together with (5.9), (5.14), (5.16), and (5.17) yields (5.7). □

Proof of Theorem 4.3. For sufficiently large n there exist $(\breve{k}, \breve{\lambda}) \in K_n \times \Lambda_n$ and a constant $B_1 > 0$ such that $\breve{k} = p$ and

$$B_1 \left(\frac{(\ln n_1)^2}{n_1} \right)^{\frac{2p}{2p+d}} J_p^2(m)^{-\frac{2p}{2p+d}} \leq \breve{\lambda} \leq 2 \cdot B_1 \left(\frac{(\ln n_1)^2}{n_1} \right)^{\frac{2p}{2p+d}} J_p^2(m)^{-\frac{2p}{2p+d}}.$$

This implies $\breve{\lambda} \to 0$ $(n \to \infty)$, $n\,\breve{\lambda} \to 0$ $(n \to \infty)$, and

$$\left(\frac{(\ln n_{\mathtt{l}})^2}{n_{\mathtt{l}}}\right)^{\frac{2p}{2p+d}} \breve{\lambda}^{-1} \to 0 \quad (n \to \infty).$$

Therefore, we can conclude from (5.7) in the proof of Theorem 4.2 and (4.8) with a suitable large chosen constant $B_2 > 0$

$$\mathbf{P}\left[\int_{\mathbb{R}^d} \left|m_{n_{\mathtt{l}},(\breve{k},\breve{\lambda})}(x) - m(x)\right|^2 \mu(dx) \right.$$
$$\left. > \frac{B_2}{n_{\mathtt{l}}}\sum_{i=1}^{n_{\mathtt{l}}} \left|Y_i^\star - \bar{Y}_i\right|^2 + B_2\left(\frac{(\ln n_{\mathtt{l}})^2}{n_{\mathtt{l}}}\right)^{\frac{2p}{2p+d}} J_p^2(m)^{\frac{d}{2p+d}}\right] \to 0 \quad (n \to \infty). \tag{5.19}$$

Now note that because of $n_{\mathtt{l}} = \lceil \frac{n}{2} \rceil$ and $n_{\mathtt{t}} = n - n_{\mathtt{l}} = \lfloor \frac{n}{2} \rfloor$

$$\frac{1}{n_{\mathtt{l}}} \le \frac{1}{n_{\mathtt{t}}} \le \frac{3}{n} \tag{5.20}$$

for all $n \in \mathbb{N}$ with $n \ge 2$. Furthermore, $\frac{n}{2} \le n_{\mathtt{l}} \le n$ implies

$$\left(\frac{(\ln n_{\mathtt{l}})^2}{n_{\mathtt{l}}}\right)^{\frac{2p}{2p+d}} J_p^2(m)^{\frac{d}{2p+d}} \le \left(2\,\frac{(\ln n)^2}{n}\right)^{\frac{2p}{2p+d}} J_p^2(m)^{\frac{d}{2p+d}} =: t_n. \tag{5.21}$$

Let $B_3 > 0$ be a suitably large chosen constant (see below). Then one can conclude from (5.20) and (5.21)

$$\mathbf{P}\left[\int_{\mathbb{R}^d} |m_n(x) - m(x)|^2 \mu(dx) > 3\,\frac{54\,B_2 + B_3}{n}\sum_{i=1}^{n}\left|Y_i^\star - \bar{Y}_i\right|^2 + (54\,B_2 + B_3)\,t_n\right]$$
$$\le \mathbf{P}\left[\int_{\mathbb{R}^d} |m_n(x) - m(x)|^2 \mu(dx) - 54 \min_{(k,\lambda)\in K_n\times\Lambda_n} \int_{\mathbb{R}^d} \left|m_{n_{\mathtt{l}},(k,\lambda)}(x) - m(x)\right|^2 \mu(dx) \right.$$
$$\left. > \frac{B_3}{n_{\mathtt{t}}}\sum_{i=n_{\mathtt{l}}+1}^{n}\left|Y_i^\star - \bar{Y}_i\right|^2 + B_3\,t_n\right]$$
$$+\mathbf{P}\left[\min_{(k,\lambda)\in K_n\times\Lambda_n} \int_{\mathbb{R}^d} \left|m_{n_{\mathtt{l}},(k,\lambda)}(x) - m(x)\right|^2 \mu(dx) \right.$$
$$\left. > \frac{B_2}{n_{\mathtt{l}}}\sum_{i=1}^{n_{\mathtt{l}}}\left|Y_i^\star - \bar{Y}_i\right|^2 + B_2\left(\frac{(\ln n_{\mathtt{l}})^2}{n_{\mathtt{l}}}\right)^{\frac{2p}{2p+d}} J_p^2(m)^{\frac{d}{2p+d}}\right].$$

The assertion of Theorem 4.3 follows from this together with (5.19) and the following lemma.

Lemma 5.2 *Let $l_1, l_2 \in \mathbb{R}$ with $l_1 \le l_2$ and $d, n \in \mathbb{N}$ with $n \ge 2$. Set $n_{\mathtt{l}} := \lceil \frac{n}{2} \rceil$ and $n_{\mathtt{t}} := n - n_{\mathtt{l}}$. Let the set of parameters $K_n \times \Lambda_n$ be defined by (2.17) and (2.18), the data $\bar{\mathcal{D}}_n$ by (2.14), and $m : [0,1]^d \to [\,l_1, l_2\,]$. For all $(k,\lambda) \in K_n \times \Lambda_n$ set $m_{n_{\mathtt{l}},(k,\lambda)}(\cdot) :=$*

$T_{[l_1,l_2]}\,\tilde m_{n_1,(k,\lambda)}(\cdot)$, where $\tilde m_{n_1,(k,\lambda)}$ is given by (4.2). Define the estimate m_n via $m_n(\cdot) := m_{n_1,(\bar k,\bar\lambda)}(\cdot)$, with

$$(\bar k,\bar\lambda) := \operatorname*{arg\,min}_{(k,\lambda)\in K_n\times\Lambda_n}\left(\frac{1}{n_{\mathtt{t}}}\sum_{i=n_1+1}^{n}|m_{n_1,(k,\lambda)}(X_i)-\bar Y_i|^2\right).$$

Set

$$H_n := \int_{\mathbb{R}^d}|m_n(x)-m(x)|^2\,\mu(dx) - 54\min_{(k,\lambda)\in K_n\times\Lambda_n}\int_{\mathbb{R}^d}\left|m_{n_1,(k,\lambda)}(x)-m(x)\right|^2\mu(dx).$$

Then there exits three constants $b_{10}, b_{11}, b_{12} > 0$, which only depend on l_1 and l_2, such that we have for all $t > 0$,

$$\begin{aligned} q_n(t) &:= \mathbf{P}\left[H_n > \frac{b_{10}}{n_{\mathtt{t}}}\sum_{i=n_1+1}^{n}\left|Y_i^\star-\bar Y_i\right|^2 + b_{10}\,t\right] \\ &\le 2\,|K_n\times\Lambda_n|\left(2\,\exp\left(-b_{11}\,n\,t\right)+\frac{\exp\left(-b_{12}\,n\,t\right)}{1-\exp\left(-b_{12}\,n\,t\right)}\right). \end{aligned} \tag{5.22}$$

Especially, we have for $t_n := \left(2\,\frac{(\ln n)^2}{n}\right)^{\frac{2p}{2p+d}} J_p^2(m)^{\frac{d}{2p+d}}$

$$q_n\left(t_n\right)\to 0 \quad (n\to\infty). \tag{5.23}$$

Proof of Lemma 5.2. For all $(k,\lambda)\in K_n\times\Lambda_n$ set

$$\nu_{k,\lambda} := \mathbf{E}\left[g_{n_1,(k,\lambda)}(X)\middle|\,\bar{\mathcal{D}}_{n_1}\right],$$

where $g_{n_1,(k,\lambda)}(x) := \left|m_{n_1,(k,\lambda)}(x)-m(x)\right|^2$ $(x\in[0,1]^d)$ and

$$\bar{\mathcal{D}}_{n_1} := \left\{\left(X_1,\bar Y_1\right),\ldots,\left(X_{n_1},\bar Y_{n_1}\right)\right\}.$$

One can conclude from (5.20) and Lemma A.2 for all $t>0$

$$\begin{aligned} q_n(t) \le\; & \mathbf{P}\left[\int_{\mathbb{R}^d} g_{n_1,(\bar k,\bar\lambda)}(x)\,\mu(dx) - \frac{2}{n_{\mathtt{t}}}\sum_{i=n_1+1}^{n} g_{n_1,(\bar k,\bar\lambda)}(X_i) > t\right] \\ & + \sum_{(k,\lambda)\in K_n\times\Lambda_n}\mathbf{P}\left[2\,\frac{1}{n_{\mathtt{t}}}\sum_{i=n_1+1}^{n} g_{n_1,(k,\lambda)}(X_i) - 3\int_{\mathbb{R}^d} g_{n_1,(k,\lambda)}(x)\,\mu(dx) > t\right] \\ & + \mathbf{P}\left[\frac{1}{n_{\mathtt{t}}}\sum_{i=n_1+1}^{n} g_{n_1,(\bar k,\bar\lambda)}(X_i) - 18\min_{(k,\lambda)\in K_n\times\Lambda_n}\frac{1}{n_{\mathtt{t}}}\sum_{i=n_1+1}^{n} g_{n_1,(k,\lambda)}(X_i)\right. \\ & \qquad\qquad\left. > t + \frac{512}{n_{\mathtt{t}}}\sum_{i=n_1+1}^{n}\left|Y_i^\star-\bar Y_i\right|^2\right] \\ & \le 2\sum_{(k,\lambda)\in K_n\times\Lambda_n}\mathbf{P}\left[\left|\nu_{k,\lambda}-\frac{1}{n_{\mathtt{t}}}\sum_{i=n_1+1}^{n} g_{n_1,(k,\lambda)}(X_i)\right| > \frac{t}{2}+\frac{\nu_{k,\lambda}}{2}\right] \end{aligned}$$

$$+\,2\;|K_n\times\Lambda_n|\,\frac{\exp\left(-b_7\,n\,t\right)}{1-\exp\left(-b_7\,n\,t\right)}, \tag{5.24}$$

with a constant $b_7>0$, which only depends on l_1 and l_2. In order to bound the first term on the right hand side of (5.24), we first note that for all $(k,\lambda)\in K_n\times\Lambda_n$

$$(l_2-l_1)^2\cdot\nu_{k,\lambda}\geq\;\mathbf{E}\left[\,(g_{n_1,(k,\lambda)}(X))^2\middle|\,\bar{\mathcal{D}}_{n_1}\right]\geq\;\mathbf{Var}\left[\,g_{n_1,(k,\lambda)}(X)\middle|\,\bar{\mathcal{D}}_{n_1}\right]=:\sigma^2_{(k,\lambda)}.$$

This together with Bernstein's inequality and (5.20) yields for all $t>0$

$$\begin{aligned}
&\sum_{(k,\lambda)\in K_n\times\Lambda_n}\mathbf{P}\left[\left|\nu_{k,\lambda}-\frac{1}{n_{\mathtt{t}}}\sum_{i=n_1+1}^{n}g_{n_1,(k,\lambda)}(X_i)\right|>\frac{t}{2}+\frac{\nu_{k,\lambda}}{2}\middle|\,\bar{\mathcal{D}}_{n_1}\right]\\
&\leq\;|K_n\times\Lambda_n|\max_{(k,\lambda)\in K_n\times\Lambda_n}\mathbf{P}\left[\left|\nu_{k,\lambda}-\frac{1}{n_t}\sum_{i=n_1+1}^{n}g_{n_1,(k,\lambda)}(X_i)\right|>\frac{t}{2}+\frac{\sigma^2_{(k,\lambda)}}{2\cdot(l_2-l_1)^2}\middle|\,\bar{\mathcal{D}}_{n_1}\right]\\
&\leq\;2\;|K_n\times\Lambda_n|\max_{(k,\lambda)\in K_n\times\Lambda_n}\exp\left(-\frac{n_{\mathtt{t}}\left(\frac{t}{2}+\frac{\sigma^2_{(k,\lambda)}}{2\cdot(l_2-l_1)^2}\right)^2}{2\sigma^2_{(k,\lambda)}+\frac{2\cdot(l_2-l_1)^2}{3}\left(\frac{t}{2}+\frac{\sigma^2_{(k,\lambda)}}{2\cdot(l_2-l_1)^2}\right)}\right)\\
&\leq\;2\;|K_n\times\Lambda_n|\max_{(k,\lambda)\in K_n\times\Lambda_n}\exp\left(-\frac{n_{\mathtt{t}}\left(\frac{t}{2}+\frac{\sigma^2_{(k,\lambda)}}{2\cdot(l_2-l_1)^2}\right)}{\frac{14}{3}\cdot(l_2-l_1)^2}\right)\\
&\leq\;2\;|K_n\times\Lambda_n|\exp\left(-\frac{n\,t}{28\cdot(l_2-l_1)^2}\right)
\end{aligned}\tag{5.25}$$

Now (5.24) and (5.25) imply (5.22). Set $B_1:=\min\{\frac{1}{28\cdot(l_2-l_1)^2},b_7\}$. Then one can conclude for sufficiently large n

$$\max\left\{\exp\left(-\frac{b_{10}}{3}\,n\,t_n\right),\exp\left(-\frac{b_7}{3}\,n\,t_n\right)\right\}\leq\exp\left(-B_9\,n\,t_n\right)\leq\frac{1}{n^2}, \tag{5.26}$$

since

$$B_1\,n\,t_n=B_1\,n\left(2\,\frac{(\ln n)^2}{n}\right)^{\frac{2p}{2p+d}}J_p^2(m)^{\frac{d}{2p+d}}\geq 2\,\ln n.$$

Now, (5.23) follows from (5.22), (5.26) and $|K_n\times\Lambda_n|\leq(\ln n)^2\,n$. □

This completes the proof of Theorem 4.3. □

6 Proof of the main results

Proof of Theorems 3.1 – 3.3. First, we shall prove the following lemma

Lemma 6.1 *Let $(Y, C), (Y_1, C_1), \ldots, (Y_n, C_n)$ be i.i.d. $\mathbb{R}_+ \times \mathbb{R}_+$-valued random vectors with Y and C independent. Let $\alpha \in \mathbb{R}$ and let Y_i^* and $\hat{Y}_i$ $(i = 1, \ldots, n)$ be defined by (2.10) and (2.13), where $G(t) = \mathbf{P}[C > t]$ $(t \in \mathbb{R})$. Set $F(t) = \mathbf{P}[Y > t]$ $(t \in \mathbb{R})$ and $\tau_F = \sup\{t : F(t) > 0\}$. Assume G is continuous and $G(\tau_F) > 0$. Then the following two results hold:*

1. $\max_{i=1,\ldots,n} |Y_i^* - \hat{Y}_i|^2 \to 0 \quad (n \to \infty) \quad$ *a.s.*
2. *Let $\gamma \in (0, 1)$. If*
$$-\int_0^{\tau_F} F(t)^{\frac{-\gamma}{2-\gamma}} dG(t) < \infty \tag{6.1}$$
then there exists a constant $b_3 \geq 0$ such that
$$\limsup_{n\to\infty} n^\gamma \max_{i=1,\ldots,n} |Y_i^* - \hat{Y}_i|^2 \leq b_3 \quad a.s. \tag{6.2}$$

Proof of Lemma 6.1. First, we note that $G(\tau_F) > 0$ implies $0 \leq Z_i = \min\{Y_i, C_i\} \leq \tau_F < \infty$ a.s. for all $i \in \{1, \ldots, n\}$. Using this, (2.10), (2.13) and the definitions of G and G_n, one can conclude with probability 1

$$\begin{aligned}
&\max_{i=1,\ldots,n} |Y_i^* - \hat{Y}_i|^2 \\
&= \max_{i=1,\ldots,n} \left| (1+\alpha) \int_0^{Z_i} \left(\frac{1}{G(t)} - \frac{1}{G_n(t)} \right) dt - \alpha \delta_i Z_i \left(\frac{1}{G(Z_i)} - \frac{1}{G_n(Z_i)} \right) \right|^2 \\
&\leq 2 \max_{i=1,\ldots,n} \left((1+|\alpha|)^2 \left| \int_0^{Z_i} \left(\frac{1}{G(t)} - \frac{1}{G_n(t)} \right) dt \right|^2 + \alpha^2 \left| \delta_i Z_i \left(\frac{1}{G(Z_i)} - \frac{1}{G_n(Z_i)} \right) \right|^2 \right) \\
&\leq 4 (1+|\alpha|)^2 \max_{i=1,\ldots,n} \left(Z_i \sup_{0 \leq t \leq Z_i} \left| \frac{1}{G(t)} - \frac{1}{G_n(t)} \right| \right)^2 \\
&\leq \frac{4 (1+|\alpha|)^2 \tau_F^2}{G(\tau_F)^2 G_n(\tau_F)^2} \left(\sup_{0 \leq t \leq \tau_F} |G(t) - G_n(t)| \right)^2.
\end{aligned} \tag{6.3}$$

Corollary 1.3 in Stute and Wang (1993), $G(\tau_F) > 0$, and

$$\begin{aligned}
\mathbf{P}\left[\limsup_{n\to\infty} \frac{1}{G_n(\tau_F)} > \frac{2}{G(\tau_F)} \right] &= \mathbf{P}\left[\limsup_{n\to\infty} (G(\tau_F) - G_n(\tau_F)) > \frac{G(\tau_F)}{2} \right] \\
&\leq \mathbf{P}\left[\limsup_{n\to\infty} \sup_{0 \leq t \leq \tau_F} |G(t) - G_n(t)| > \frac{G(\tau_F)}{2} \right]
\end{aligned}$$

imply

$$\limsup_{n\to\infty} \frac{1}{G_n(\tau_F)^2} \leq \frac{4}{G(\tau_F)^2} \quad \text{a.s.} \tag{6.4}$$

From (6.3) and (6.4) one can conclude for all $\gamma \in [0, 1)$ that with probability 1

$$\limsup_{n\to\infty} n^\gamma \max_{i=1,\ldots,n} |Y_i^* - \hat{Y}_i|^2$$

$$\leq \quad \frac{4\,(1+|\alpha|)^2\,\tau_F^2}{G(\tau_F)^2} \limsup_{n\to\infty} \frac{1}{G_n(\tau_F)^2} \left(n^{\frac{\gamma}{2}} \sup_{0\leq t\leq \tau_F} |G(t)-G_n(t)| \right)^2$$

$$\leq \quad \frac{16\,(1+|\alpha|)^2\,\tau_F^2}{G(\tau_F)^4} \limsup_{n\to\infty} \left(n^{\frac{\gamma}{2}} \sup_{0\leq t\leq \tau_F} |G(t)-G_n(t)| \right)^2 . \tag{6.5}$$

For $\gamma = 0$, this together with Corollary 1.3 in Stute and Wang (1993) implies the assertion of part 1 of Lemma 6.1. Now, let $\gamma \in (0,1)$ and assume that (6.1) holds. Then part 2 of Lemma 6.1 follows from (6.5) and Theorem 2.1 in Chen and Lo (1997). □

Now, we start with the proof of our main results. First notice that (2.10), $G(L) = \mathbf{P}\,[\,C > L\,] > 0$, and $0 \leq Z = \min\{Y, C\} \leq L$ a.s. imply

$$|Y^*| \leq (1+2|\alpha|)\,\frac{L}{G(L)} =: L^\star < \infty \quad \text{a.s.} \tag{6.6}$$

For all $(k,\lambda) \in K_n \times \Lambda_n$, let $m_{n_1,(k,\lambda)}$ be defined by (2.19). Since $|m(X)| \leq L \leq L^\star$ a.s., we have for all $k \in \mathbb{N}$ with $2k > d$ and all $\lambda_n > 0$

$$\int_{\mathbb{R}^d} |m_{n,(k,\lambda_n)}(x) - m(x)|^2 \mu(dx) \leq \int_{\mathbb{R}^d} |T_{[-L^\star,L^\star]}\tilde{m}_{n,(k,\lambda_n)}(x) - m(x)|^2 \mu(dx). \tag{6.7}$$

Therefore, one can conclude from 6.1 (where we set $\gamma := \frac{2p}{2p+d}$ and note that, since $2p > d$, $\frac{2p}{2p+d} \in \left(\frac{1}{2}, 1\right)$) and (6.6), that the assertions of Theorem 3.1 and Theorem 3.2 follow from Theorem 4.1 and Theorem 4.2, respectively.

In order to prove Theorem 3.3, we first recall the definitions

$$(\hat{k},\hat{\lambda}) = \operatorname*{arg\,min}_{(k,\lambda)\in K_n\times\Lambda_n} \left(\frac{1}{n_{\mathtt{t}}} \sum_{i=n_1+1}^{n} |T_{[0,L]}\tilde{m}_{n_1,(k,\lambda)}(X_i) - \hat{Y}_i|^2 \right)$$

and

$$(\bar{k},\bar{\lambda}) = \operatorname*{arg\,min}_{(k,\lambda)\in K_n\times\Lambda_n} \left(\frac{1}{n_{\mathtt{t}}} \sum_{i=n_1+1}^{n} |T_{[-L^\star,L^\star]}\tilde{m}_{n_1,(k,\lambda)}(X_i) - \hat{Y}_i|^2 \right).$$

Now, let $B_1 > 0$ be a suitably large chosen constant. Set

$$t_n := \left(\frac{(\ln n)^2}{n} \right)^{\frac{2p}{2p+d}} J_p^2(m)^{\frac{d}{2p+d}}$$

and

$$H_n := \int_{\mathbb{R}^d} |m_{n_1,(\hat{k},\hat{\lambda})}(x) - m(x)|^2 \mu(dx) - 54 \min_{(k,\lambda)\in K_n\times\Lambda_n} \int_{\mathbb{R}^d} |m_{n_1,(k,\lambda)}(x) - m(x)|^2 \mu(dx).$$

Then one can conclude with (5.20) and (6.7)

$$\mathbf{P}\left[\int_{\mathbb{R}^d} |m_{n_1,(\hat{k},\hat{\lambda})}(x) - m(x)|^2 \mu(dx) > 109\,B_1\,t_n \right]$$

$$
\begin{aligned}
&\leq \mathbf{P}\left[H_n > B_1\, t_n + \frac{B_1}{n_{\mathtt{t}}}\sum_{i=n_1+1}^{n}|Y_i^* - \hat{Y}_i|^2\right] \\
&\quad + \mathbf{P}\left[\int_{\mathbb{R}^d}|m_{n_1,(\bar{k},\bar{\lambda})}(x) - m(x)|^2\mu(dx) > 2\, B_1\, t_n - \frac{B_1}{54 n_{\mathtt{t}}}\sum_{i=n_1+1}^{n}|Y_i^* - \hat{Y}_i|^2\right] \\
&\leq \mathbf{P}\left[H_n > B_1\, t_n + \frac{B_1}{n_{\mathtt{t}}}\sum_{i=n_1+1}^{n}|Y_i^* - \hat{Y}_i|^2\right] \\
&\quad + \mathbf{P}\left[\int_{\mathbb{R}^d}|T_{[-L^\star,L^\star]}\tilde{m}_{n_1,(\bar{k},\bar{\lambda})}(x) - m(x)|^2\mu(dx) > B_1\, t_n + \frac{B_1}{n}\sum_{i=1}^{n}|Y_i^* - \hat{Y}_i|^2\right] \\
&\quad + \mathbf{P}\left[\frac{1}{n}\sum_{i=1}^{n}|Y_i^* - \hat{Y}_i|^2 > \frac{t_n}{2}\right]
\end{aligned}
$$

This together with Theorem 4.3, Lemma 5.2, and Lemma 6.1 implies the assertion of Theorem 3.3. □

A Results for fixed design regression

Below we formulate and prove two auxiliary results which are used in the proofs of Theorem 4.2 and 4.3, in a fixed design regression model. Let $x_1, \dots, x_n \in [0,1]^d$ be arbitrary, but fixed. Let $m : [0,1]^d \to \mathbb{R}^d$ and assume for all $i = 1, \dots, n$

$$
Y_i^\star = m(x_i) + U_i, \tag{A.1}
$$

where $U_1, \dots, U_n$ are independent random variables with expectation zero. Then the following two results hold

Lemma A.1 *Let $n, d \in \mathbb{N}$, $\lambda_n > 0$, $L^\star \geq 1$, and $b_4 > 0$. Set $l := L^\star + b_4$ and let $p \in \mathbb{N}$ with $2p > d$ be arbitrary. Let $Y_1^\star, \dots, Y_n^\star$ be given by (A.1) with $|Y_i^\star| \leq L^\star$ a.s. and $|m(x_i)| \leq L^\star$ for all $i \in \{1, \dots, n\}$. Let $\bar{Y}_1, \dots, \bar{Y}_n$ be arbitrary real-valued random variables and define the estimates $\tilde{m}_{n,(p,\lambda_n)}$ and $m^\star_{n,(p,\lambda_n)}$ by*

$$
\tilde{m}_{n,(p,\lambda_n)}(\cdot) := \underset{f \in W_p([0,1]^d)}{\arg\min}\left(\frac{1}{n}\sum_{i=1}^{n}|f(x_i) - \bar{Y}_i|^2 + \lambda_n J_p^2(f)\right), \tag{A.2}
$$

where $W_p([0,1]^d)$ and $J_p^2(f)$ ($f \in W_p([0,1]^d)$ are given by (2.3) and (2.5), and

$$
m^\star_{n,(p,\lambda_n)}(\cdot) := T_{[-l,l]}\tilde{m}_{n,(p,\lambda_n)}(\cdot). \tag{A.3}
$$

Assume $m \in W_p([0,1]^d)$ with $J_p^2(m) < \infty$ and set

$$
V_{1,n} := \frac{1}{n}\sum_{i=1}^{n}|m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i)|^2 + \lambda_n J_p^2(\tilde{m}_{n,(p,\lambda_n)}) - 2\lambda_n J_p^2(m) - \frac{64}{n}\sum_{i=1}^{n}|Y_i^\star - \bar{Y}_i|^2.
$$

Then there exist constants $b_5, b_6 > 0$ *which only depend on* $L^\star$, *such that for any* $t_n > 0$ *with*

$$t_n \quad \to 0 \quad (n \to \infty), \tag{A.4}$$

$$\frac{n\, t_n}{\ln n} \quad \to \infty \quad (n \to \infty), \tag{A.5}$$

and

$$\frac{n\, t_n}{\ln n} \lambda_n^{\frac{d}{2p}} \to \infty \quad (n \to \infty), \tag{A.6}$$

we have for all $t \geq t_n$ *and all sufficiently large* n

$$\mathbf{P}\left[V_{1,n} > t,\, \frac{1}{n}\sum_{i=1}^{n} |m^\star_{n,(p,\lambda_n)}(x_i) - \bar{Y}_i|^2 \leq \frac{1}{n}\sum_{i=1}^{n} |\tilde{m}_{n,(p,\lambda_n)}(x_i) - \bar{Y}_i|^2 \right] \leq b_5 \exp\left(-b_6 n t\right).$$

Proof of Lemma A.1. First, we notice that

$$U_i^2 = |Y_i^\star - m(x_i)|^2 \leq 4(L^\star)^2 \quad \text{a.s.} \tag{A.7}$$

for all $i \in \{1, \dots, n\}$. Set

$$H_n := \frac{1}{n}\sum_{i=1}^{n} |m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i)|^2 + \lambda_n J_p^2\left(\tilde{m}_{n,(p,\lambda_n)}\right).$$

By an application of Lemma B.1 in combination with (A.7), we have for all $t > 0$

$$\begin{aligned} &\mathbf{P}\left[V_{1,n} > t,\, \frac{1}{n}\sum_{i=1}^{n} |m^\star_{n,(p,\lambda_n)}(x_i) - \bar{Y}_i|^2 \leq \frac{1}{n}\sum_{i=1}^{n} |\tilde{m}_{n,(p,\lambda_n)}(x_i) - \bar{Y}_i|^2 \right] \\ \leq\;& \mathbf{P}\left[t < H_n \leq \frac{8}{n}\sum_{i=1}^{n} (m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i))\, U_i,\, \frac{1}{n}\sum_{i=1}^{n} U_i^2 \leq 4(L^\star)^2 \right] \\ =:\;& q_{1,n}. \end{aligned} \tag{A.8}$$

In order to derive an upper bound on $q_{1,n}$, we notice that (A.7) together with the Cauchy–Schwarz inequality yields

$$\frac{8}{n}\sum_{i=1}^{n} (m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i))\, U_i \leq 16 L^\star \sqrt{\frac{1}{n}\sum_{i=1}^{n} |m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i)|^2} \quad \text{a.s.}$$

Therefore one can conclude that inside of $q_{1,n}$

$$\begin{aligned} H_n \;&\leq\; \left(\sqrt{\frac{1}{n}\sum_{i=1}^{n} |m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i)|^2} + \frac{\lambda_n J_p^2(\tilde{m}_{n,(p,\lambda_n)})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} |m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i)|^2}} \right)^2 \\ &=\; \left(\frac{H_n}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} |m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i)|^2}} \right)^2 \leq 256 (L^\star)^2. \end{aligned}$$

For arbitrary $t > 0$ set

$$\bar{\jmath}_{min} := \min\left\{j \in \mathbb{N} : 2^j t \geq 256(L^\star)^2\right\}.$$

By an application of the Peeling-technique (cf. (5.10)), we can conclude from (A.8) for all $t > 0$

$$\begin{aligned}
&\mathbf{P}\left[V_{1,n} > t,\ \frac{1}{n}\sum_{i=1}^n |m^\star_{n,(p,\lambda_n)}(x_i) - \bar{Y}_i|^2 \leq \frac{1}{n}\sum_{i=1}^n |\tilde{m}_{n,(p,\lambda_n)}(x_i) - \bar{Y}_i|^2\right] \\
\leq\ & \sum_{j=1}^{\bar{\jmath}_{min}} \mathbf{P}\left[\frac{2^j t}{2} < H_n \leq 2^j t, H_n \leq \frac{8}{n}\sum_{i=1}^n (m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i))\, U_i, \frac{1}{n}\sum_{i=1}^n U_i^2 \leq 4(L^\star)^2\right] \\
\leq\ & \sum_{j=1}^{\bar{\jmath}_{min}} \mathbf{P}\left[H_n \leq 2^j t, \frac{1}{n}\sum_{i=1}^n (m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i))\, U_i > \frac{2^j t}{16}, \frac{1}{n}\sum_{i=1}^n U_i^2 \leq 4(L^\star)^2\right] \\
\leq\ & \sum_{j=1}^{\bar{\jmath}_{min}} \mathbf{P}\left[\sup_{g \in \mathcal{G}_{2^j t/\lambda_n}} \left|\frac{1}{n}\sum_{i=1}^n U_i\, g(x_i)\right| \geq \frac{2^j t}{16}, \frac{1}{n}\sum_{i=1}^n U_i^2 \leq 4(L^\star)^2\right] \\
=:\ & \sum_{j=1}^{\bar{\jmath}_{min}} q_{2,n,j},
\end{aligned} \tag{A.9}$$

where for all $j \in \{1, \ldots, \bar{\jmath}_{min}\}$

$$\mathcal{G}_{2^j t/\lambda_n} := \left\{f - m : f \in \mathcal{F}_{2^j t/\lambda_n}, \frac{1}{n}\sum_{i=1}^n |f(x_i) - m(x_i)|^2 \leq 2^j t\right\}$$

with

$$\mathcal{F}_{2^j t/\lambda_n} := \left\{T_{[-l,l]} f : f \in W_p([0,1]^d), J_p^2(f) \leq \frac{2^j t}{\lambda_n}\right\}.$$

Similar to the proof of Theorem 4.2 (vide (5.13)), one can conclude from Lemma B.2 and (A.5) for all $t \geq t_n$, all $j \in \{1, \ldots, \bar{\jmath}_{min}\}$, and all sufficiently large n

$$\int_0^{\sqrt{2^j t}} \sqrt{\ln \mathcal{N}_2\left(s, \mathcal{G}_{2^j t/\lambda_n}, x_1^n\right)}\, ds \leq \sqrt{n} 2^j t \left(b_8 \sqrt{\frac{\ln n}{n t_n} \lambda_n^{-\frac{d}{2p}}} + b_9 \sqrt{\frac{\ln n}{n t_n}}\right),$$

with some constants $b_8, b_9 > 0$ which only depend on p and d. This together with (A.5), (A.6), and Corollary 8.3 in van de Geer (2000) (set there $K := 2L^\star$, $\sigma_0 := 2\sqrt{2}L^\star$, $\delta := \frac{2^j t}{16}$, $\sigma := 2L^\star$, and $R := \sqrt{2^j t}$) implies for all $t \geq t_n$ and all sufficiently large n

$$\sum_{j=1}^{\bar{\jmath}_{min}} q_{2,n,j} \leq \sum_{j=1}^{\bar{\jmath}_{min}} B_1 \cdot \exp\left(-B_2 n 2^j t\right) \leq B_1 \frac{\exp(-B_2 n t)}{1 - \exp(-B_2 n t)} \leq 2B_1 \exp(-B_2 n t), \tag{A.10}$$

where $B_1, B_2 > 0$ are two constants which only depend on $L^\star$. The assertion of Lemma A.1 follows from (A.9) and (A.10). □

Lemma A.2 *Let $d, n_1, n_{\mathtt{t}} \in \mathbb{N}$ with $n_1 + n_{\mathtt{t}} =: n$ and $l_1, l_2 \in \mathbb{R}$ with $l_1 \leq l_2$. Let $Y_1^\star, \ldots, Y_n^\star$ be given by (A.1) with $Y_i^\star, m(x_i) \in [l_1, l_2]$ a.s. for all $i \in \{1, \ldots, n\}$. Let $\bar{Y}_1, \ldots, \bar{Y}_n$ be arbitrary real-valued random variables and let $K \times \Lambda$ be a (finite) set of parameters with $K \subseteq \mathbb{N}_d$, where $\mathbb{N}_1 := \mathbb{N}$ for $d = 1$ and $\mathbb{N}_d := \mathbb{N} \setminus \{1, \ldots, \lfloor \frac{d}{2} \rfloor\}$ for $d > 1$, and $\Lambda \subseteq \mathbb{R}_+ \setminus \{0\}$. For each $(k, \lambda) \in K \times \Lambda$ define the estimates $\tilde{m}_{n_1,(k,\lambda)}$ and $m_{n_1,(k,\lambda)}$ by*

$$\tilde{m}_{n_1,(k,\lambda)}(\cdot) := \operatorname*{arg\,min}_{f \in W_k([0,1]^d)} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} |f(x_i) - \bar{Y}_i|^2 + \lambda J_k^2(f) \right),$$

where $W_k([0,1]^d)$ and $J_k^2(f)$ ($f \in W_k([0,1]^d)$ are given by (2.3) and (2.5), and

$$m_{n_1,(k,\lambda)}(\cdot) := T_{[l_1,l_2]} \tilde{m}_{n_1,(k,\lambda)}(\cdot).$$

Now, let

$$m_n(\cdot) := \operatorname*{arg\,min}_{f \in \mathcal{F}_{K \times \Lambda}} \left(\frac{1}{n_{\mathtt{t}}} \sum_{i=n_1+1}^{n} |f(x_i) - \bar{Y}_i|^2 \right)$$

where

$$\mathcal{F}_{K \times \Lambda} := \left\{ m_{n_1,(k,\lambda)} : (k, \lambda) \in K \times \Lambda \right\}.$$

Set

$$V_{2,n} := \frac{1}{n_{\mathtt{t}}} \sum_{i=n_1+1}^{n} |m_n(x_i) - m(x_i)|^2 - 18 \min_{(k,\lambda) \in K \times \Lambda} \frac{1}{n_{\mathtt{t}}} \sum_{i=n_1+1}^{n} \left| m_{n_1,(k,\lambda)}(x_i) - m(x_i) \right|^2.$$

Then there exists a constant $b_7 > 0$ which depends only on l_1 and l_2, such that for all $t > 0$

$$\mathbf{P}\left[V_{2,n} > t + \frac{512}{n_{\mathtt{t}}} \sum_{i=n_1+1}^{n} \left| Y_i^\star - \bar{Y}_i \right|^2 \right] \leq 2 \, |K \times \Lambda| \, \frac{\exp(-b_7 n_{\mathtt{t}} t)}{1 - \exp(-b_7 n_{\mathtt{t}} t)}$$

Proof of Lemma A.2. Set

$$m_n^*(\cdot) := \operatorname*{arg\,min}_{f \in \mathcal{F}_{K \times \Lambda}} \frac{1}{n_{\mathtt{t}}} \sum_{i=n_1+1}^{n} |f(x_i) - m(x_i)|^2.$$

By Lemma 1 in Kohler (2002), one can conclude for arbitrary $t > 0$

$$\begin{aligned}
&\mathbf{P}\left[V_{2,n} > t + \frac{512}{n_{\mathtt{t}}} \sum_{i=n_1+1}^{n} \left| Y_i^\star - \bar{Y}_i \right|^2 \right] \\
\leq \; & \mathbf{P}\Bigg[\frac{t}{2} < \frac{1}{n_{\mathtt{t}}} \sum_{i=n_1+1}^{n} |m_n(x_i) - m_n^*(x_i)|^2 \\
& \qquad \leq \frac{16}{n_{\mathtt{t}}} \sum_{i=n_1+1}^{n} (m_n(x_i) - m_n^*(x_i)) (Y_i^\star - m(x_i)) \Bigg] \\
\leq \; & |K \times \Lambda| \max_{(k,\lambda) \in K \times \Lambda} \mathbf{P}\left[\frac{t}{2} < H_{1,n,k,\lambda} \leq 16 \, H_{2,n,k,\lambda} \right],
\end{aligned}$$

where

$$H_{1,n,k,\lambda} := \frac{1}{n_{\mathtt{t}}} \sum_{i=n_1+1}^{n} \left| m_{n_1,(k,\lambda)}(x_i) - m_n^*(x_i) \right|^2$$

and

$$H_{2,n,k,\lambda} := \frac{1}{n_{\mathtt{t}}} \sum_{i=n_1+1}^{n} \left(m_{n_1,(k,\lambda)}(x_i) - m_n^*(x_i) \right) \left(Y_i^{\star} - m(x_i) \right).$$

An application of the Peeling-technique (cf. (5.10)) yields

$$\begin{aligned} & \mathbf{P}\left[V_{2,n} > t + \frac{512}{n_{\mathtt{t}}} \sum_{i=n_1+1}^{n} \left| Y_i^{\star} - \bar{Y}_i \right|^2 \right] \\ \leq \;& |K \times \Lambda| \max_{(k,\lambda) \in K \times \Lambda} \sum_{s=0}^{\infty} \mathbf{P}\left[\frac{2^s t}{2} < H_{1,n,k,\lambda} \leq 2^s t, H_{1,n,k,\lambda} \leq 16\, H_{2,n,k,\lambda} \right] \\ \leq \;& |K \times \Lambda| \max_{(k,\lambda) \in K \times \Lambda} \sum_{s=0}^{\infty} \mathbf{P}\left[H_{1,n,k,\lambda} \leq 2^s t, H_{2,n,k,\lambda} > \frac{2^s t}{32} \right]. \end{aligned}$$

Set $b_7 := \frac{2}{(2(l_2 - l_1))^2 32^2}$. By Hoeffding's inequality, we have for all $t > 0$

$$\begin{aligned} \mathbf{P}\left[V_{2,n} > t + \frac{512}{n_{\mathtt{t}}} \sum_{i=n_1+1}^{n} \left| Y_i^{\star} - \bar{Y}_i \right|^2 \right] \;&\leq\; |K \times \Lambda| \sum_{s=0}^{\infty} 2 \exp\left(-b_7\, t\, n_{\mathtt{t}}\, 2^s \right) \\ &\leq\; 2\, |K \times \Lambda| \, \frac{\exp\left(-b_7 n_{\mathtt{t}} t \right)}{1 - \exp\left(-b_7 n_{\mathtt{t}} t \right)}. \end{aligned}$$

□

B Two deterministic lemmata

This section contains two deterministic lemmata which are used in the proofs of Theorem 4.2 and Lemma A.1.

Lemma B.1 *Let $l \geq 0$, $t > 0$, $d \in \mathbb{N}$, $x_1, \ldots, x_n \in [0,1]^d$, $y_1^{\star}, \bar{y}_1, \ldots, y_n^{\star}, \bar{y}_n \in \mathbb{R}$, and $m : [0,1]^d \to \mathbb{R}$. Let $p \in \mathbb{N}$ with $2p > d$ be arbitrary, $\lambda_n > 0$, and let the estimates $\tilde{m}_{n,(p,\lambda_n)}$ and $m^{\star}_{n,(p,\lambda_n)}$ be defined by*

$$\tilde{m}_{n,(p,\lambda_n)}(\cdot) := \operatorname*{arg\,min}_{f \in W_p([0,1]^d)} \left(\frac{1}{n} \sum_{i=1}^{n} |f(x_i) - \bar{y}_i|^2 + \lambda_n J_p^2(f) \right), \tag{B.1}$$

where $W_p([0,1]^d)$ and $J_p^2(f)$ ($f \in W_p([0,1]^d)$ are given by (2.3) and (2.5), and

$$m^{\star}_{n,(p,\lambda_n)}(\cdot) := T_{[-l,l]} \tilde{m}_{n,(p,\lambda_n)}(\cdot). \tag{B.2}$$

Set

$$V_{3,n} := \frac{1}{n} \sum_{i=1}^{n} |m^{\star}_{n,(p,\lambda_n)}(x_i) - m(x_i)|^2 + \lambda_n J_p^2(\tilde{m}_{n,(p,\lambda_n)}).$$

If $m \in W_p([0,1]^d)$ *with* $J_p^2(m) < \infty$,

$$V_{3,n} > t + \frac{64}{n} \sum_{i=1}^{n} |y_i^\star - \bar{y}_i|^2 + 2\lambda_n J_p^2(m), \tag{B.3}$$

and

$$\frac{1}{n} \sum_{i=1}^{n} |m^\star_{n,(p,\lambda_n)}(x_i) - \bar{y}_i|^2 \le \frac{1}{n} \sum_{i=1}^{n} |\tilde{m}_{n,(p,\lambda_n)}(x_i) - \bar{y}_i|^2, \tag{B.4}$$

then we have

$$V_{3,n} \le \frac{8}{n} \sum_{i=1}^{n} \left(m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i) \right) (y_i^\star - m(x_i)). \tag{B.5}$$

Proof of Lemma B.1. Assume $m \in W_p([0,1]^d)$ with $J_p^2(m) < \infty$. Then we have by Definition (B.1) and inequality (B.4)

$$\begin{aligned} V_{3,n} &= \frac{1}{n} \sum_{i=1}^{n} |m^\star_{n,(p,\lambda_n)}(x_i) - \bar{y}_i|^2 + \lambda_n J_p^2(\tilde{m}_{n,(p,\lambda_n)}) - \frac{1}{n} \sum_{i=1}^{n} |m(x_i) - \bar{y}_i|^2 \\ &\quad + \frac{2}{n} \sum_{i=1}^{n} (m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i)) (\bar{y}_i - m(x_i)) \\ &\le \lambda_n J_p^2(m) + \frac{2}{n} \sum_{i=1}^{n} (m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i)) (\bar{y}_i - m(x_i)) \end{aligned} \tag{B.6}$$

If

$$\frac{2}{n} \sum_{i=1}^{n} (m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i)) (\bar{y}_i - m(x_i)) < \lambda_n J_p^2(m),$$

then we can conclude from (B.3) and (B.6)

$$t + 2\lambda_n J_p^2(m) < V_{3,n} < 2\lambda_n J_p^2(m)$$

in contradiction to $t > 0$. Therefore, we have shown that

$$V_{3,n} \le \frac{4}{n} \sum_{i=1}^{n} (m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i)) (\bar{y}_i - y_i^\star) + \frac{4}{n} \sum_{i=1}^{n} (m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i)) (y_i^\star - m(x_i)) \tag{B.7}$$

Now assume that the second term of the right hand side of (B.7) is smaller than the first one. Then we can conclude from (B.7) by an application of the Cauchy–Schwarz inequality

$$V_{3,n} \le 8 \sqrt{\frac{1}{n} \sum_{i=1}^{n} |m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i)|^2} \cdot \sqrt{\frac{1}{n} \sum_{i=1}^{n} |y_i^\star - \bar{y}_i|^2}. \tag{B.8}$$

If

$$\frac{1}{n} \sum_{i=1}^{n} |m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i)|^2 \ne 0,$$

then (B.8) together with (B.3) implies

$$
\begin{aligned}
& t + \frac{64}{n} \sum_{i=1}^{n} |y_i^\star - \bar{y}_i|^2 \\
< & \left(\sqrt{\frac{1}{n} \sum_{i=1}^{n} |m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i)|^2} + \frac{\lambda_n J_p^2(\tilde{m}_{n,(p,\lambda_n)})}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} |m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i)|^2}} \right)^2 \\
\leq & \frac{64}{n} \sum_{i=1}^{n} |y_i^\star - \bar{y}_i|^2
\end{aligned}
$$

in contradiction to $t > 0$. And if

$$\frac{1}{n} \sum_{i=1}^{n} |m^\star_{n,(p,\lambda_n)}(x_i) - m(x_i)|^2 = 0,$$

we can conclude from (B.3) and (B.8) $t < \lambda_n J_p^2(\tilde{m}_{n,(p,\lambda_n)}) < 0$. From this together with (B.7), the assertion (B.5) of Lemma B.1 follows. □

Lemma B.2 *Let $l, b > 0$ and $p, d, n \in \mathbb{N}$ with $2p > d$ and $n > 1$. Set*

$$F_b := \left\{ T_{[-l,l]} f : f \in W_p([0,1]^d), J_p^2(f) \leq b \right\},$$

where $W_p([0,1]^d)$ and $J_p^2(f)$ $(f \in W_p([0,1]^d)$ are given by (2.3) and (2.5). Then there exist constants $b_8, b_9 > 0$ which only depend on p and d, such that for all $\zeta \geq \frac{l^2}{n}$ and all $x_1, \dots, x_n \in [0,1]^d$

$$\int_0^{\sqrt{\zeta}} \sqrt{\ln \mathcal{N}_2(s, \mathcal{F}_b, x_1^n)}\, ds \leq b_8 \left(\frac{b}{\zeta}\right)^{\frac{d}{4p}} \sqrt{\zeta}\sqrt{\ln n} + b_9 \sqrt{\zeta}\sqrt{\ln n}. \qquad \text{(B.9)}$$

Proof of Lemma B.2. For any $\zeta > 0$ and all $x_1, \dots, x_n \in [0,1]^d$ set

$$I_\zeta := \int_0^{\sqrt{\zeta}} \sqrt{\ln \mathcal{N}_2(s, \mathcal{F}_b, x_1^n)}\, ds.$$

Lemma 3 in Kohler et al. (2002) implies that there exist two constants $B_1, B_2 > 0$ which only depend on p and d, such that for all $\zeta > 0$ and all $x_1, \dots, x_n \in [0,1]^d$

$$I_\zeta \leq B_1 b^{\frac{d}{4p}} \int_0^{\sqrt{\zeta}} s^{-\frac{d}{2p}} \sqrt{\ln\left(\frac{64 l^2 e n}{s^2}\right)}\, ds + B_2 \int_0^{\sqrt{\zeta}} \sqrt{\ln\left(\frac{64 l^2 e n}{s^2}\right)}\, ds. \qquad \text{(B.10)}$$

Substituting $t := \frac{\sqrt{\zeta}}{s}$ and applying Hölders inequality, one can conclude for all $\zeta > 0$ and all $x_1, \dots, x_n \in [0,1]^d$ from (B.10)

$$I_\zeta \leq B_1 b^{\frac{d}{4p}} \zeta^{\frac{1}{2} - \frac{d}{4p}} \sqrt{\int_1^\infty t^{\frac{d}{2p} - 2} dt \cdot \int_1^\infty t^{\frac{d}{2p} - 2} \ln\left(\frac{64 l^2 e n}{\zeta}\, t^2\right) dt}$$

$$+B_2\sqrt{\zeta}\sqrt{\int_1^\infty t^{-2}dt \cdot \int_1^\infty t^{-2}\ln\left(\frac{64l^2en}{\zeta}t^2\right)dt}$$

$$= B_3\left(\frac{b}{\zeta}\right)^{\frac{d}{4p}}\sqrt{\zeta}\sqrt{\ln\left(B_4\frac{l^2n}{\zeta}\right)} + B_2\sqrt{\zeta}\sqrt{\ln\left(B_5\frac{l^2n}{\zeta}\right)} \tag{B.11}$$

with the constants $B_3 := \frac{B_1}{1-\frac{d}{2p}}$, $B_4 := 64e^{1+\frac{2}{1-\frac{d}{2p}}}$, and $B_5 := 64e^3$. Finally, for all $\zeta \geq \frac{l^2}{n}$ and all $x_1, \ldots, x_n \in [0,1]^d$, (B.11) implies

$$I_\zeta \leq B_3\left(\frac{b}{\zeta}\right)^{\frac{d}{4p}}\sqrt{\zeta}\sqrt{2\left(2+\ln B_4\right)}\sqrt{\ln n} + B_2\sqrt{\zeta}\sqrt{2\left(2+\ln B_5\right)}\sqrt{\ln n}.$$

□

Acknowledgements. The authors thank M. Kohler and K. Máthé for many helpful discussions and advice which improved this article. This work is part of a so called Verbundforschungsprojekt between the Fraunhofer Institut für Grenzflächen- und Bioverfahrenstechnik (IGB) Stuttgart, the Robert-Bosch-Krankenhaus (RBK) Stuttgart, the Institut für Zellbiologie und Immunologie (IZI), the Institut für Stochastik und Anwendungen (ISA) (both Universität Stuttgart) as well as the Interfakultäres Institut für Zellbiologie (IZT) (Universität Tübingen) and was supported by the Land Baden-Württemberg (BW PtJ-BIO/24-720.431-1-1/D31).

References

[1] Andersen, P. K., Borgan, Ø., Gill, R.D., and Keiding, N. (1993). *Statistical models based on counting processes.* Springer Series in Statistics, Springer-Verlag, New York.

[2] Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical Report, University of California, Berkeley.

[3] Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrica* **66** 429–436.

[4] Carbonez, A. (1992). *Nonparametric functional estimation under random censoring and a new semiparametric model of random censorship.* PhD thesis, Katholieke Universiteit Leuven.

[5] Chen, K. and Lo, S.-H. (1997). On the rate of uniform convergence of the product-limit estimator: strong and weak laws. *Ann. Statist.* **25** 1050–1087.

[6] Cox, D.R. and Oakes, D. (1984). *Analysis of survival data.* Chapman & Hall, London.

[7] Dippon, J. (2004). Kernel and support vector methods for censored survival data with covariates. Submitted for publication.

[8] Duchon, J. (1976). Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. (French) *R.A.I.R.O. Analyse Numérique* **10** 5–12.

[9] Fan, J., and Gijbels, I. (1994). Censored regression: local linear approximations and their applications. *J. Amer. Statist. Assoc.* **89** 560–570.

[10] Fan, J., and Gijbels, I. (1996). *Local polynomial modelling and its applications.* Chapman & Hall, London.

[11] Flemington, T.R. and Harrington, D.P. (1991). *Counting processes and survival analysis.* Wiley, New York.

[12] Györfi, L., Kohler, M., Krzyżak, A., and Walk, H. (2002). *A distribution-free theory of nonparametric regression.* Springer Series in Statistics, Springer-Verlag, New York.

[13] Huang, J. Z. and Stone, C. J. (1998). The $\mathcal{L}_2$ rate of convergence for event history regression with time-dependent covariates. *Scand. J. Statist.* **25** 603–620.

[14] James, I. R., and Smith, P. J. (1984). Consistency results for linear regression with censored data. *Ann. Statist.* **12** 590–600.

[15] Kaplan, E. L., and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

[16] Kohler, M. (2000). Inequalities for uniform deviations of averages from expectations with applications to nonparametric regression. *J. Stat. Plann. and Inf.* **89** 1–23.

[17] Kohler, M. (2002). Nonparametric regression with additional measurement errors in the dependent variable. Preprint 2002–14, Fakultät Mathematik, Universität Stuttgart.

[18] Kohler, M. and Krzyżak, A. (2001). Nonparametric regression estimation using penalized least squares. *IEEE Trans. Inform. Theory* **47** 3054–3058.

[19] Kohler, M., Krzyżak, A., and Schäfer D. (2002). Application of structural risk minimization to multivariate smoothing spline regression estimates. *Bernoulli* **8** 475–489.

[20] Kooperberg, C., Stone, C. J., Troung, Y. K., (1995a). Hazard regression. *J. Amer. Statist. Assoc.* **90** 78–94.

[21] Kooperberg, C., Stone, C. J., Troung, Y. K., (1995b). The $\mathcal{L}_2$ rate of convergence for hazard regression *Scand. J. Statist.* **22** 143–157.

[22] Koul, H., Susarla, V., and Van Ryzin, J. (1981). Regression analysis with randomly right–censored data *Ann. Statist.* **9** 1276–1288.

[23] Lai, T. L., and Ying, Z. (1991). Large sample theory of a modified Buckley–James estimator for regression analysis with censored data. *Ann. Statist.* **19** 1370–1402.

[24] Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrica* **74** 301–309.

[25] Miller, R. G. (1976). Least squares regression with censored data. *Biometrika* **63** 449–464.

[26] Park, J. (2004). Optimal global rate of convergence in nonparametric regression with left-truncated and right-censored data. *J. Multivariate Anal.* **89** 70–86.

[27] Pintér, M. (2001). Consistency results in nonparametric regression and classification. PhD thesis, Budapest University of Technology and Economics.

[28] Pollard, D. (1984). *Convergence of stochastic processes.* Springer Series in Statistics, Springer-Verlag, New York.

[29] Rudin, W. (1987). *Real and complex analysis.* Mc Graw-Hill series in higher mathematics, McGraw-Hill Book Co., New York.

[30] Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18** 303–328.

[31] Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

[32] Van de Geer, S., 2000. *Empirical processes in M-estimation.* Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.

[33] Wahba, G. (1990). *Spline models for observational data.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia.

[34] Zheng, Z. (1987). A class of estimators of the parameters in linear regression with censored data. *Acta Math. Appl. Sin. Engl. Ser.* **3**, 231–241.

[35] Zucker, D. M. and Karr, A.F. (1990). Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *Ann. Statist.* **18** 329–353.

Jürgen Dippon
Pfaffenwaldring 57
70569 Stuttgart
Germany
E-Mail: dippon@mathematik.uni-stuttgart.de
WWW: http://www.isa.uni-stuttgart.de/LstStoch/Dippon

Stefan Winter
Pfaffenwaldring 57
70569 Stuttgart
Germany
E-Mail: wintersn@mathematik.uni-stuttgart.de
WWW: http://www.isa.uni-stuttgart.de/LstStoch/Winter

Erschienene Preprints ab Nummer 2004/001

Komplette Liste: `http://www.mathematik.uni-stuttgart.de/preprints`

2004/001 *Walk, H.:* Strong Laws of Large Numbers by Elementary Tauberian Arguments.

2004/002 *Hesse, C.H., Meister, A.:* Optimal Iterative Density Deconvolution: Upper and Lower Bounds.

2004/003 *Meister, A.:* On the effect of misspecifying the error density in a deconvolution problem.

2004/004 *Meister, A.:* Deconvolution Density Estimation with a Testing Procedure for the Error Distribution.

2004/005 *Efendiev, M.A., Wendland, W.L.:* On the degree of quasiruled Fredholm maps and nonlinear Riemann-Hilbert problems.

2004/006 *Dippon, J., Walk, H.:* An elementary analytical proof of Blackwell's renewal theorem.

2004/007 *Mielke, A., Zelik, S.:* Infinite-dimensional hyperbolic sets and spatio-temporal chaos in reaction-diffusion systems in $\mathbb{R}^n$.

2004/008 *Exner, P., Linde, H., Weidl T.:* Lieb-Thirring inequalities for geometrically induced bound states.

2004/009 *Ekholm, T., Kovarik, H.:* Stability of the magnetic Schrödinger operator in a waveguide.

2004/010 *Dillen, F., Kühnel, W.:* Total curvature of complete submanifolds of Euclidean space.

2004/011 *Afendikov, A.L., Mielke, A.:* Dynamical properties of spatially non-decaying 2D Navier-Stokes flows with Kolmogorov forcing in an infinite strip.

2004/012 *Pöschel, J.:* Hill's potentials in weighted Sobolev spaces and their spectral gaps.

2004/013 *Dippon, J., Walk, H.:* Simplified analytical proof of Blackwell's renewal theorem.

2004/014 *Kühnel, W.:* Tight embeddings of simply connected 4-manifolds.

2004/015 *Kühnel, W., Steller, M.:* On closed Weingarten surfaces.

2004/016 *Leitner, F.:* On pseudo-Hermitian Einstein spaces.

2004/017 *Förster, C., Östensson, J.:* Lieb-Thirring Inequalities for Higher Order Differential Operators.

2005/001 *Mielke, A.; Schmid, F.:* Vortex pinning in super-conductivity as a rate-independent model

2005/002 *Kimmerle, W.; Luca, F., Raggi-Cárdenas, A.G.:* Irreducible Components of the Burnside Ring

2005/003 *Höfert, C.; Kimmerle, W.:* On Torsion Units of Integral Group Rings of Groups of Small Order

2005/004 *Röhrl, N.:* A Least Squares Functional for Solving Inverse Sturm-Liouville Problems

2005/005 *Borisov, D.; Ekholm, T; Kovarik, H.:* Spectrum of the Magnetic Schrödinger Operator in a Waveguide with Combined Boundary Conditions

2005/006 *Zelik, S.:* Spatially nondecaying solutions of 2D Navier-Stokes equation in a strip

2005/007 *Meister, A.:* Deconvolving compactly supported densities

2005/008 *Förster, C., Weidl, T.:* Trapped modes for an elastic strip with perturbation of the material properties

2006/001 *Dippon, J., Schiemert, D.:* Stochastic differential equations driven by Gaussian processes with dependent increments

2006/002 *Lesky, P.A.:* Orthogonale Polynomlösungen von Differenzengleichungen vierter Ordnung

2006/003 *Dippon, J., Schiemert, D.:* Option Pricing in a Black-Scholes Market with Memory

2006/004 *Banchoff, T., Kühnel, W.:* Tight polyhedral models of isoparametric families, and PL-taut submanifolds

2006/005 *Walk, H.:* A universal strong law of large numbers for conditional expectations via nearest neighbors

2006/006 *Dippon, J., Winter, S.:* Smoothing spline regression estimates for randomly right censored data

2006/007 *Walk, H.:* Almost sure Cesàro and Euler summability of sequences of dependent random variables

2006/008 *Meister, A.:* Optimal convergence rates for density estimation from grouped data

2006/009 *Förster, C.:* Trapped modes for the elastic plate with a perturbation of Young's modulus