Multivariate boundary regression models

Leonie Selk[*]   Charles Tillier[†]   Orlando Marigliano[‡]

April 22, 2020

Abstract

In this work, we consider a nonparametric regression model with one-sided errors, multivariate covariates and regression function in a general Hölder class. We work under the assumption of regularly varying independent and identically distributed errors that are also independent of the design points. Following Drees, Neumeyer and Selk (2019), we estimate the regression function, i.e, the upper boundary curve, via minimization of the local integral of a polynomial approximation lying above the data points. The main purpose of this paper is to show the uniform consistency and to provide the rates of convergence of such estimators for both multivariate random covariates and multivariate deterministic design points. To demonstrate the performance of the estimators, the small sample behavior is investigated in a simulation study in dimension two and three.

Key words: Frontier estimation, boundary models, multivariate analysis, nonparametric regression, uniform rates of convergence, extreme value theory, regular variation, local polynomial approximation

1 Introduction

We consider nonparametric regression models with one-sided errors that take the general form

$$Y_i = g(X_i) + \varepsilon_i, \quad i = 1, \ldots, n \tag{1.1}$$

[*]University of Hamburg, Department of Mathematics, Bundesstrasse 55, 20146 Hamburg, Germany. Email: leonie.selk@math.uni-hamburg.de

[†]University of Versailles-Saint-Quentin-en-Yvelines, Laboratoire de Mathématiques de Versailles, 45 avenue des états-unis, 78000 Versailles, France. Email: charles.tillier@uvsq.fr

[‡]Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig, Germany. Email: Orlando.Marigliano@mis.mpg.de

where Y_i is the response variable, X_i is the covariate, g is the unknown regression function corresponding to the upper boundary curve and ε_i is a nonpositive random error term. The statistical issue of such a model lies hence on the frontier estimation, in other words, on the estimation of g based on the observations (X_i, Y_i), $i = 1, \ldots, n$ where n is the sample size of the available data.

Models described in Equation (1.1) are also called *boundary regression models* (BRM) and have received an increasing attention in the last past years. BRM are closely related to production frontier models (PFM). Both share the objective of the estimation of the frontier - the boundary curve - and contribute to the same applications. PFM appeared first in the field of data envelopment analysis (DEA). Introduced in 1950's with the seminal contribution of Farrell (1957), DEA answers the need of developing nonparametric methods to evaluate productivity and to assess efficiency of a system. A production unit is technically efficient if it produces the maximum output which is technically feasible for given inputs, or uses minimal inputs for the production of a given level of output. Relaxing the fundamental hypothesis of convex hull body of the data in DEA models, Deprins, Simar and Tulkens (1984) later introduced free disposal hull (FDH) models; see also Lovell et al. (1994) for further developments. DEA and its extensions are now recognized as powerful quantitative, analytical tools for measuring and evaluating the performance of a system and have countless applications, for instance in social sciences, health-care evaluation systems, banking sectors, supply chain management, energy system assessment or when analysing the performance of public services like hospitals and education. We refer to the books Cooper, Seiford and Zhu (2011) and Ramanathan (2003) for an comprehensive treatment of such methods and an exhaustive development of the applications. On the other hand, stochastic frontier analysis (SFA) originally formulated independently by Aigner, Lovell and Schmidt (1977) and Meeusen and van Den Broeck (1977) offers an interesting alternative with parametric estimations of the frontier; see also the books of Kumbhakar and Lovell (2003) and Cornwell and Schmidt (2008) for more recent references.

There is vast literature in the theory of production frontier models dealing with the question of estimating the boundary curve. Numerous parametric and non parametric techniques have been proposed, for instance using extreme-value based estimators (see for instance Girard and Jacob (2003), de Haan and Resnick (1994), Hall, Nussbaum and Stern (1997), Menneteau (2008), Girard and Jacob (2004), Gardes (2002) and Gijbels and Peng (2000)), projections techniques (Jacob and Suquet (1995)), kernel based estimators (Girard, Guillou and Stupfler (2013), Girard and Jacob (2008)), maximum likelihood based estimators (see Kumbhakar et al. (2007) and Simar and Zelenyuk (2011)). Some estimators need the boundary curve to be monotone (see e.g. Daouia and Simar (2005), Daouia, Noh and Park (2016) and Gijbels et al. (1999)).

In contrast to the aforementioned methods, we concentrate in this paper on an alternative approach that consists in approximating the regression function g locally by a polynomial lying above the data points. Polynomial estimators in frontier estimation have been widely studied in the literature and are employed in several works, for instance in Hall, Park and Stern (1998), Hall and Park (2004), Girard, Iouditski and Nazin (2005), Knight (2001) and in Hall and Van Keilegom (2009) (local-linear estimator); see also the literature using piecewise polynomials (e.g. Korostelëv and Tsybakov (1993), Korostelëv, Silmar and Tsybakov (1995) and Härdle, Park and Tsybakov (1995)).

While consistency of estimators based on local constant approximations (local polynomial estimator of order 0) of the boundary curve may be attained under mild assumptions of the boundary curve - such as continuity of g in Neumeyer, Selk and Tillier (2019) - rates of convergence need stronger regularity assumption on g. A classical framework is to assume that g belongs to some Hölder class of order β. This assumption essentially means that g is β-times differentiable and moreover a Lipschitz condition holds for the β derivative. Further assumptions are also needed on the error distribution such as regular variation, meaning that the distribution of the errors has polynomial tails. These two assumptions are common in the context of boundary models and are met in several papers, see for instance Meister and Reiß (2013), Jirak, Meister and Reiß (2014), Müller and Wefelmeyer (2010), Drees, Neumeyer and Selk (2019), Girard et al. (2013), Härdle et al. (1995) and Hall and Van Dellegom (2009); see also the book of de Haan and Ferriera (2006) for the applications and the motivation of regularly varying errors.

In the context of nonparametric regression models with an Hölder boundary curve and regularly varying nonpositive errors, Jirak et al. (2014) suggested an adaptive estimator for the boundary curve g using a local polynomial estimation based on local extreme value statistics. An adaptative procedure - a fully data-driven estimation procedure - is constructed by applying a nested version of Lepski method which shows no loss in the convergence rates with respect to the general L_q-risk. Drees et al. (2019) estimated the regression function similarly to Jirak et al. (2014) via minimization of the local integral of a polynomial approximation in the context of equidistant design points. By showing uniform rates of convergence for the regression estimator they proposed distribution-free tests of error distributions (goodness-of-fit) where the test statistics are based on empirical processes of residuals. They also discussed asymptotically distribution-free hypotheses tests for independence of the error distribution from the points of measurement and for monotonicity of the boundary function as well.

More often than not, fixed (deterministic) covariates and more specifically equidistant fixed design points are considered. In contrast in the paper at hand, we investigate the two cases of random and deterministic covariates which are both of particular interest. Deterministic covariates are often used in real-life applications when time is involved in the

data set. For instance Jirak et al. (2014) studied the monthly sunspot observations and the annual best running times of 1500 meters; see also the plentiful applications in energy and environmental performance analysis provided in the review of Mardani et al. (2018). Besides, deterministic design is met accross a number of papers in regression models, see for instance Brown and Low (1996), Meister and Reiß (2013) and the references within. The case of random covariates is obviously the most relevant and appears in essence in many applications in boundary models, among other, in non-life insurance mathematics and financial risk modelling when analyzing optimality of portfolios and efficiency strategies; see also the extensive literature on modern portoflio theory (e.g. the recent books of Francis and Kim (2013) and Goetzmann et al. (2014)).

In opposition to kernel estimation methods, all the literature cited above concerning the asymptotic study of local polynomial approximations of the boundary curve deal with univariate covariates and as far as we are concerned the topic of providing rates of convergence of such estimators based on a multivariate sample is new and challenging. Beyond the theoretical interest this generalisation raises, we think that extending to the multivariate setup is also interesting from an application point of view. In light of this motivation, the main aim of this paper is to extend the results of Drees et al. (2019) to the multivariate case with arbitrary covariate. That is, under the main assumptions of regular variation of the nonpositive errors and β-Hölder class of the boundary curve, we aim at showing uniform consistency and providing rates of convergence of an estimator based on the minimisation of the local integral of a polynomial lying above the data points for both *multivariate random covariates* and *multivariate deterministic design points*.

The remaining part of the manuscript is organized as follows. In section 2 the model is explained, while in section 3 the estimation procedure is described. In section 4 we show uniform consistency and provide rates of convergence of the estimator of the regression function for both random and deterministic multivariate covariates. Section 5 is dedicated to a simulation study to investigate the small sample behavior in dimension two and three. Proofs are postponed to the appendix.

Notation and shortcuts

Notation: $\mathcal{B}$ stands for Borelian sets; $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ are the floor and ceiling functions respectively; $\langle x \rangle$ means the largest natural number that is strictly smaller than x; $\bar{F} = 1 - F$ denotes the survival function associated to a cdf F; $\| \cdot \|_\infty$ denotes the supremum norm; $X_1 \stackrel{d}{=} X_2$ means that two random variables X_1, X_2 share the same distribution; $a_n \underset{n \to \infty}{\sim} b_n$ holds if $\lim_{n \to \infty} a_n / b_n = 1$ for two sequences $(a_n)_{n \geq 1}$ and $(b_n)_{n \geq 1}$ of nonnegative numbers. Generally, vectors are highlighted in bold writing. For vectors $\mathbf{x} \in \mathbb{R}^q$ let $x^{(r)}$ denote the r-th component of $\mathbf{x}$ for $r = 1, \ldots, q$. By $\|\mathbf{x}\|$ we mean the maximum norm that is $\|\mathbf{x}\| := \max_{r \in \{1, \ldots, q\}} |x^{(r)}|$.

For multivariate polynomials we use the multiindex notation where for a vector $\mathbf{x} \in \mathbb{R}^q$ and a multiindex $\mathbf{j} = (j_1, \ldots, j_q) \in \mathbb{N}_0^q$ we define $|\mathbf{j}| := j_1 + \ldots + j_q$, $\mathbf{j}! := j_1! \cdot \ldots \cdot j_q!$ and $\mathbf{x}^{\mathbf{j}} := x^{(1)^{j_1}} \cdot \ldots \cdot x^{(q)^{j_q}}$.

Shortcuts: *cdf* stands for cumulative distribution function and *iid* for independent and identically distributed.

2   The model

We focus on nonparametric boundary regression models with univariate observations Y_i and multivariate covariates X_i of the form

$$Y_i = g(X_i) + \varepsilon_i, \quad i = 1, \ldots, n \tag{2.1}$$

where the errors ε_i are iid non-positive univariate random variables that are independent of the X_i and n stands for the sample size. The unknown regression function g thus corresponds to the upper boundary curve.

Independence of errors ε_i and covariates X_i is a typical assumption in regression models and is met among others in Müller and Wefelmeyer (2010), Meister and Reiß (2013), Reiß and Selk (2017) and Drees et al. (2019). Such assumption is crucial and often needed in the framework of frontier models when using statistical methods such as bootstrap procedures; see Wilson (2003) and Simar and Wilson (1998). In the case this hypothesis does not hold, one may consider parametric transformations e.g. exponential, Box-Cox (Box and Cox (1964)), or sinh-arcsinh transformations (Jones and Pewsey (2009)) of the response variable Y_i in order to retreive the framework of independence between errors and covariates; see Neumeyer et al. (2019) for the univariate case and also Linton, Sperlich and Van Keilegom (2008).

In the paper at hand, we investigate both cases of random and fixed design points. The former means that the covariates X_i are q-dimensional random variables while the latter assumes that X_i are deterministically spread over $[0, 1]^q$. Without loss of generality, we work for ease of reading with design points lying on $[0, 1]^q$ but results extend effortless to any cartesian product of one-dimensional closed intervals.

2.1   The random design case

In the random design case we consider the nonparametric boundary regression model with independent and identically distributed observations $(\mathbf{X}_i, Y_i)$, $i = 1, \ldots, n$ defined by

$$Y_i = g(\mathbf{X}_i) + \varepsilon_i \tag{2.2}$$

corresponding to model (2.1), where the design points $\mathbf{X}_i$ are multivariate random covariates distributed on $[0, 1]^q$ that fulfill assumption (K4). The errors ε_i are assumed to be iid non-positive random variables that satisfy (K2). The precise statements of assumptions (K2) and (K4) are given in Section 3.1.

2.2 The fixed design case

In the fixed design case we consider a triangular array of independent observations $Y_{i,n}$ and deterministic design points $\mathbf{x}_{i,n}$ in $[0, 1]^q$ for $i = 1, \ldots, n$. Thus we conduct the nonparametric boundary regression model

$$Y_{i,n} = g(\mathbf{x}_{i,n}) + \varepsilon_{i,n}, \tag{2.3}$$

corresponding to model (2.1) with errors $\varepsilon_{i,n}$ that are univariate non-positive independent and identically distributed random variables that satisfy assumption (K2).

We allow for fixed equidistant as well as fixed nonequidistant design. In the first case we consider $\mathbf{x}_{1,n}, \ldots, \mathbf{x}_{n,n}$ that form a grid

$$\begin{pmatrix} n^{-\frac{1}{q}} \\ n^{-\frac{1}{q}} \\ \vdots \\ n^{-\frac{1}{q}} \end{pmatrix}, \begin{pmatrix} 2n^{-\frac{1}{q}} \\ n^{-\frac{1}{q}} \\ \vdots \\ n^{-\frac{1}{q}} \end{pmatrix}, \ldots, \begin{pmatrix} 1 \\ \vdots \\ 1 \\ (n^{\frac{1}{q}} - 1)n^{-\frac{1}{q}} \end{pmatrix}, \begin{pmatrix} 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}$$

where we assume that $n^{\frac{1}{q}}$ is an integer. Note that when $q = 1$ the univariate equidistant design simplifies to $\mathbf{x}_{i,n} = i/n$ for $i = 1, \ldots, n$. In the second case the points are not necessarily equidistant, but we assume that they are even enough distributed on $[0, 1]^q$, see Assumption (K4') below.

3 Estimating the regression function

To estimate the boundary curve, we use an estimator that locally approximates the regression function g by a polynomial lying above the data points; see Theorem 2.2 in Drees, Neumeyer and Selk (2019) for further details in the univariate case.

3.1 The random design case

For $\mathbf{x} \in [0, 1]^q$, we consider the regression function estimator $\hat{g}$ defined as

$$\hat{g}_n(\mathbf{x}) := \hat{g}(\mathbf{x}) := p(\mathbf{x}) \tag{3.1}$$

where p is a multivariate polynomial of total degree $\beta^* \in \mathbb{N}_0$ and minimizes the local integral

$$\int p(\mathbf{t})I\{\|\mathbf{t} - \mathbf{x}\| \leq h_n\}d\mathbf{t} \tag{3.2}$$

under the constraints $p(\mathbf{X}_i) \geq Y_i$ for $\|\mathbf{X}_i - \mathbf{x}\| \leq h_n$.

Remark 3.1 *The polynomial p is the solution to the linear optimization problem*

$$\begin{aligned} \text{minimize} \quad & v^T p \\ \text{subject to} \quad & Ap \geq y, \end{aligned} \tag{3.3}$$

where p is represented by its vector of coefficients, v is the vector representing the linear functional $\int_{\|\mathbf{t}-\mathbf{x}\| \leq h_n}$, the matrix A is the multivariate Vandermonde matrix whose i-th row has as its entries all the monomials of degree at most β^* in the entries of $\mathbf{X}_i$, and y is the vector with $y_i = Y_i$. For the estimator $\hat{g}$ to be well-defined, it is necessary that this problem is bounded from below, that is, that the objective is bounded from below on the polytope defined by the constraints. This need not always be the case, as the example in Jirak et al. (2014) with $\beta^* = 2$ and two support points demonstrates. However, the alternate optimization problem proposed in Jirak et al. (2014) has the same problem of unboundedness. When $q = 1$, Problem (3.3) is bounded whenever we have at least $\beta^* + 1$ points. This follows from the fact that the univariate Vandermonde matrix is totally positive when the support points X_i are positive. However, for higher dimensions it is not as simple. There, the Vandermonde matrix with $\binom{q+\beta^*}{q}$ rows needs not be invertible. As of now, we believe that for $q > 1$ the boundedness of (3.3) needs to be checked on a case-by-case basis using linear optimization algorithms. By duality theory we know that (3.3) is bounded if and only if there exists a vector $g \geq 0$ with $A^T g = v$. This linear program can become very large as q and β^* grow. For instance, if $q = \beta^*$ and A has $N = \binom{2q}{q}$ rows, then the interior point algorithm from Vaidya (1989) runs in $O(N^{2.5})$ time in the worst case, which in terms of q grows faster than $4^{2.5q} \cdot q^{-1.25}$. Nevertheless, this is a theoretical worst case and the average case might be better, possibly using another algorithm. For implementations, the authors suggest using the Python module `scipy.optimize.linprog`, *which by default uses an interior point algorithm based on the MOSEK interior point optimizer by Andersen and Andersen (2000).*

Remark 3.2 *To illustrate the estimation procedure we take a look at the simplest case of $\beta^* = 0$ which results in a local constant approximation. The estimator defined in (3.1)-(3.2) then simplifies to $\hat{g}(\mathbf{x}) = \max\{Y_i \mid i = 1, \ldots, n \text{ with } \|\mathbf{X}_i - \mathbf{x}\| \leq h_n\}$. In Figure 1 an example for $q = 1$ and uniformly distributed $\mathbf{X}_i$ is shown. For each $\mathbf{x}$-value a constant function is fitted to the data in the neighborhood.*
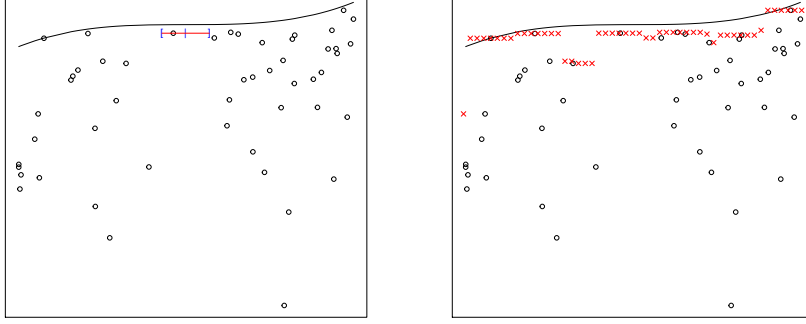
Figure 1: Scatter plot of $(\mathbf{X}_i, Y_i)$, $i = 1, \ldots, n$ together with the true regression function (black solid curve) in $q = 1$. A local constant approximation is considered. On the left hand side the blue vertical bar marks the point $(\mathbf{x}, \hat{g}(\mathbf{x}))$ for some given $\mathbf{x}$ and the blue square brackets the endpoints of the interval $[\mathbf{x} - h_n, \mathbf{x} + h_n]$. On the right hand side the red crosses indicate the estimated values $\hat{g}(\mathbf{x})$ for different values of $\mathbf{x}$.

We work under the following four assumptions (K1)-(K4).

(K1) **Regression function**: g belongs to some Hölder class of order $\beta \in (0, \infty)$ that is g is $\lfloor \beta \rfloor$-times differentiable on $[0, 1]^q$ and all partial derivatives of order $\lfloor \beta \rfloor$ satisfy

$$|D^{\mathbf{j}} g(\mathbf{t}) - D^{\mathbf{j}} g(\mathbf{x})| \leq c_g \|\mathbf{t} - \mathbf{x}\|^{\beta - \lfloor \beta \rfloor} \qquad \forall \mathbf{x}, \mathbf{t} \in [0, 1]^q, \ \forall \mathbf{j} \in \mathbb{N}_0^q \text{ with } |\mathbf{j}| = \lfloor \beta \rfloor \quad (3.4)$$

for some $c_g < \infty$ where
$$D^{\mathbf{j}} = \frac{\partial^{j_1 + \ldots + j_q}}{\partial x_1^{j_1} \ldots \partial x_q^{j_q}}.$$

(K2) **Errors distribution**: The errors ε_i are independent and identically distributed on $(-\infty, 0]$ with common cdf F that satisfies

$$\overline{F}(y) = c|y|^\alpha + r(y), \quad y < 0,$$

with $\alpha, c > 0$ and $r(y) = o(|y|^\alpha)$ when $y \nearrow 0$.

(K3) **Bandwidths**: $(h_n)_{n \in \mathbb{N}}$ is a sequence of positive bandwidths that satisfies $\lim_{n \to \infty} h_n = 0$ and $\lim_{n \to \infty} (\log n)/(n h_n^q) = 0$.

(K4) **Design points**: The covariates $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are iid random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and valued on $[0, 1]^q$ with cdf $F_{\mathbf{X}}$ and density $f_{\mathbf{X}}$ that is bounded and bounded away from zero. Besides, they are independent of the errors $\varepsilon_1, \ldots, \varepsilon_n$.

8

3.2 The fixed design case

Similarly to the random design case, for $\mathbf{x} \in [0,1]^q$, we consider the regression function estimator $\hat{g}$ defined as

$$\hat{g}_n(\mathbf{x}) := \hat{g}(\mathbf{x}) := p(\mathbf{x}) \tag{3.5}$$

where p is a polynomial of total degree $\beta^* \in \mathbb{N}_0$ and minimizes the local integral

$$\int p(\mathbf{t}) I\{\|\mathbf{t} - \mathbf{x}\| \leq h_n\} d\mathbf{t} \tag{3.6}$$

under the constraints $p(\mathbf{x}_{i,n}) \geq Y_{i,n}$ for $\|\mathbf{x}_{i,n} - \mathbf{x}\| \leq h_n$.

We work for the fixed design case under (K1)-(K2) and the following modified assumptions (K3') and (K4').

(K3') **Bandwidths**: Let $(h_n)_{n \geq 0}$ be a sequence of positive bandwidths that satisfies $\lim_{n \to \infty} h_n = 0$ and $\lim_{n \to \infty} \log(n)/d_n = 0$ with $d_n = d_n(1)$ from (K4').

(K4') **Design points**: Let $I_n \subset [0,1]^q$ be a q-dimensional interval which is the cartesian product of one-dimensional closed intervals of length dh_n with $d > 0$. We assume that at least $d_n(d)$ and at most $d_n'(d)$ design points lie in all such I_n.

Remark 3.3 *Assumptions (K3) and (K3') are common when analyzing the asymptotic behavior of such estimators of the form (3.1) or (3.5). In the equidistant fixed design framework we have $d_n(d) = nh_n^q d^q$, $d_n' = (n^{\frac{1}{q}} h_n d + 1)^q$ and (K3') equals (K3). Then for the univariate case $q = 1$, we have $d_n = nh_n d$ and assumption (K3') turns to be assumption (H1) in Drees et al. (2019); see also assumption (A4') in Neumeyer et al. (2019).*

Remark 3.4 *It is clear for assumption (K2) that the errors do not depend on the covariates in both cases of random and fixed design setups. Still, it has to be noticed that when dealing with the triangular scheme defined in (2.3), the errors depend on n too. This justifies the addition of the second index in $\varepsilon_{i,n}$. Indeed, the ith design point $\mathbf{x}_{i,n}$ may vary with the sample size; see the construction of the fixed multivariate equidistant design in Section 2.2 to be convinced.*

Remark 3.5 *The size of α in assumption (K2) is an important factor for the performance of the estimator defined in (3.1)-(3.2) and (3.5)-(3.6) respectively. Simply speaking the smaller $\alpha > 0$, the better the estimator. For $\alpha < 2$ the error distribution is irregular and in this case the rate of convergence for the estimator is faster than the typical nonparametric rate, see the considerations below Theorem 4.1. In Figure 2 we show some examples for different error distributions to highlight the effect of the size of α. To simplify the presentation we restrict the display to $q = 1$.*
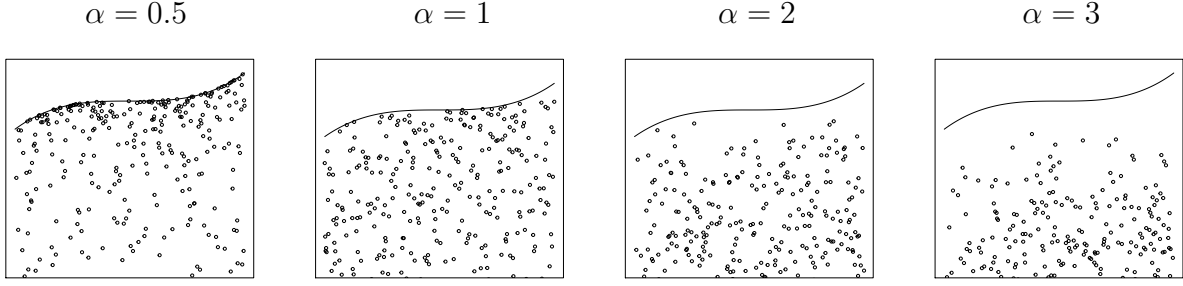
Figure 2: Scatter plots of $(\frac{i}{n}, Y_{i,n})$, $i = 1, \ldots, n$ (fixed equidistant design for $q = 1$) and the true regression function $g(\mathbf{x}) = (\mathbf{x} - 0.5)^3 + 2$. The errors are Weibull distributed, s. t. $F(y) = \exp(-|y|^\alpha)I_{(-\infty,0)}(y) + I_{[0,\infty)}(y)$ for different values of α.

4 Main results

In this section we give the uniform consistency as well as the convergence rates of our estimator $\hat{g}$ separated for the two considered cases.

4.1 The random design case

In the next theorem, we provide the uniform consistency as well as the rate of convergence of the estimator of the regression function defined in (3.1)-(3.2) for the random design case.

Theorem 4.1 *Assume model (2.2) holds. If (K1) holds with $\beta \in (0, \beta^* + 1]$ and (K2)-(K4) are satisfied, then the estimator of the regression function defined in (3.1)-(3.2) is uniformly consistent on $[0, 1]^q$ and we have*

$$\sup_{\mathbf{x}\in[0,1]^q} |\hat{g}(\mathbf{x}) - g(\mathbf{x})| = O(h_n^\beta) + O_{\mathbb{P}}\left(\left(\frac{\log(n)}{nh_n^q}\right)^{1/\alpha}\right).$$

Proof: See Appendix.

Note that the deterministic part $O(h_n^\beta)$ stems from the approximation of the regression function by a polynomial whereas the random part $O_{\mathbb{P}}\left(\left(\frac{\log(n)}{nh_n^q}\right)^{1/\alpha}\right)$ results from the observational error. Balancing the two error rates by setting $h_n \approx (\log(n)/n)^{\frac{1}{\alpha\beta+q}}$ gives $\sup_{\mathbf{x}\in[0,1]^q} |\hat{g}(\mathbf{x}) - g(\mathbf{x})| = O_{\mathbb{P}}((\log(n)/n)^{\frac{\beta}{\alpha\beta+q}})$. For the case of an irregular error distribution, i. e. $\alpha \in (0, 2)$, this rate improves upon the typical optimal rate $O_{\mathbb{P}}((\log(n)/n)^{\frac{\beta}{2\beta+q}})$ for the nonparametric estimation of mean regression functions in models with regular errors.

Remark 4.2 *Theorem 4.1 extends the result of Drees et al. (2019), Theorem 2.2, to the multivariate random setting. Even in the univariate deterministic setting our result (Theorem*

10

4.3 established below) is an extension of the aforementioned Theorem since the convergence rate holds on the whole unit interval $[0, 1]$ whereas in Drees et al. (2019) the result is restricted to $[h_n, 1 - h_n]$. The proof of the error rate that stems from the observational error follows along similar lines as the proof in Drees et al. (2019) but major adaptions are needed to deal with the multivariate and random case. The proof of the deterministic error rate that is due to approximating the boundary curve g by a polynomial is based on the proof of Theorem 3.1 in Jirak et al. (2014). In the univariate equidistant fixed case that is treated in Drees et al. (2019) this Theorem can be applied directly whereas in the multivariate, possibly random case that is treated in the paper at hand the proof has to be intensely modified. Indeed, the original proof in Jirak et al. (2014) fully relies on the fundamental theorem of algebra, which states that every polynomial equation in one variable with complex coefficients has at least one complex solution. As far as we know there is no possible extension of such a result for higher dimension hence moving to multidimensional covariates requires completely different arguments. See the proof of Proposition A.1 and especially the proof of Lemma A.2 for this modification. This also gives an alternative proof of Theorem 3.1 in Jirak et al. (2014) and extends it to the multivariate and possibly random case.

4.2   The fixed design case

We give in the next theorem the uniform consistency as well as the rate of convergence of the estimator of the regression function defined in (3.5)-(3.6) for deterministic design points.

Theorem 4.3 *Assume model (2.3) holds. If (K1) holds with $\beta \in (0, \beta^* + 1]$ and (K2), (K3') and (K4') are satisfied, then the estimator of the regression function defined in (3.5)-(3.6) is uniformly consistent on $[0, 1]^q$ and we have*

$$\sup_{\mathbf{x} \in [0,1]^q} |\hat{g}(\mathbf{x}) - g(\mathbf{x})| = O(h_n^\beta) + O_{\mathbb{P}}\left(\left(\frac{\log(n)}{d_n} \right)^{1/\alpha} \right).$$

Proof: See Appendix.

When $\beta^* = 0$ the Hölder class defined in Assumption (K1) reduces to the so-called class of *β-Hölder uniformly continuous functions* with $\beta \in (0, 1]$. In this framework, the boundary curve g may be estimated by a local constant approximation

$$\hat{g}(\mathbf{x}) = \max\{Y_{i,n} | i = 1, \ldots, n \text{ with } \|\mathbf{x}_{i,n} - \mathbf{x}\| \leq h_n\}. \tag{4.1}$$

This local constant approximation based estimator has been studied in Neumeyer et al. (2019) in the univariate setup for both random and fixed design points. Under the weaker assumption of continuity of the boundary curve g, they showed the uniform consistency of

the estimator defined in (4.1) on the whole unite interval $[0, 1]$. Strengthening with the β-Hölder uniformly continuity assumption and iid regularly varying innovations, we obtain from Theorem 4.3 for the multivariate case, the uniform consistency as well as the rate of convergence for the deterministic design case on the whole unit interval $[0, 1]^q$. We sum up in the following corollary.

Corollary 4.4 *Assume (K2), (K3') and (K4') hold for model* (2.3) *with β-Hölder uniformly continuous function g. Then, the local constant approximation of the boundary curve g defined in* (4.1) *is uniformly consistent on $[0, 1]^q$ and we have*

$$\sup_{\mathbf{x} \in [0,1]^q} |\hat{g}(\mathbf{x}) - g(\mathbf{x})| = O(h_n^{\beta}) + O_{\mathbb{P}}\left(\left(\frac{\log(n)}{nh_n} \right)^{1/\alpha} \right).$$

Remark 4.5 *Corollary 4.4 extends Remark 2.5 in Drees et al. (2019) to the multivariate setup and for non necessarily equidistant design points.*

5 Simulations

To study the small sample behavior, we generate data according to the model

$$Y_i = 0.5 \cdot \sin\left(2\pi(X_i^{(1)} + \ldots + X_i^{(q)})\right) + 4(X_i^{(1)} + \ldots + X_i^{(q)}) - \varepsilon_i$$

with $\mathbf{X}_1, \ldots, \mathbf{X}_n$ iid $\sim \text{Unif}([0, 1]^q)$ and $\varepsilon_1, \ldots, \varepsilon_n$ iid $\sim \text{Exp}(1)$ for $q = 2, 3$ and sample sizes $n = 10^q, 20^q, 30^q$. Thus in the simulated model the error distribution fulfills $\alpha = 1$ and we consider different values of $\beta^* = 0, 1, 2, 3$ which means that we investigate local constant, local linear, local quadratic and local cubic approximation.

Since our estimator is computed by minimizing an integral over a polynomial the estimation procedure consists of solving a linear programming problem (compare to Remark 3.1). The bandwidth is chosen as $h_n = n^{-\frac{1}{\beta^*+1+q}}$ which corresponds (up to a log term) to the theoretically optimal bandwidth for $\alpha = 1$ that balances the two error rates, see the considerations below Theorem 4.1.

We run the same simulations with fixed equidistant design points $\mathbf{x}_{1,n}, \ldots, \mathbf{x}_{n,n}$ as described in section 2.2. The results are very similar and thus we only present the results for the random design case.

In Table 1 the results for 1000 replications are shown where we display the estimated mean squared error of our estimator $\hat{g}(0.5, 0.5)$ ($\hat{g}(0.5, 0.5, 0.5)$ respectively). It can be seen that the estimator works reasonably well and the results improve as β^* grows. The results improve as well as n grows which are both expected effects that correspond to the theoretical result in Theorem 4.1.

$q=2$	$n=10^2$	$n=20^2$	$n=30^2$	$q=3$	$n=10^3$	$n=20^3$	$n=30^3$
$\beta^*=0$	1.56	0.95	0.62	$\beta^*=0$	0.62	0.06	0.02
$\beta^*=1$	0.02	0.006	0.002	$\beta^*=1$	0.06	0.02	0.005
$\beta^*=2$	0.07	0.01	0.005	$\beta^*=2$	0.03	0.003	0.001
$\beta^*=3$	0.02	0.003	0.001	$\beta^*=3$	0.02	0.001	0.0003

Table 1: Estimated mean squared error for $\hat{g}(0.5, 0.5)$ (left-hand side) and for $\hat{g}(0.5, 0.5, 0.5)$ (right-hand side)

$q=2$	$n=10^2$	$n=20^2$	$n=30^2$	$q=3$	$n=10^3$	$n=20^3$	$n=30^3$
$\beta^*=0$	0.79	0.37	0.23	$\beta^*=0$	1.18	0.52	0.3
$\beta^*=1$	0.22	0.06	0.04	$\beta^*=1$	0.15	0.08	0.05
$\beta^*=2$	1.4	0.03	0.009	$\beta^*=2$	0.09	0.008	0.002
$\beta^*=3$	7.47	0.05	0.01	$\beta^*=3$	0.56	0.008	0.002

Table 2: Arithmetic mean of estimated mean squared errors for $\hat{g}$ on $[0,1]^2$ (left-hand side) and on $[0,1]^3$ (right-hand side)

In Figure 3 we show the true boundary curve and the estimated curve $\hat{g}$ in comparison. Since in dimension one the presentation of two curves in one plot is clearer than in higher dimensions we display a cut through the two respectively three dimensional surface of the functions. To be precise we plot the functions $x \mapsto 0.5\sin(2\pi(x+0.5)) + 4(x+0.5)$ and $x \mapsto \hat{g}(x, 0.5)$ ($x \mapsto 0.5\sin(2\pi(x+1)) + 4(x+1)$ and $x \mapsto \hat{g}(x, 0.5, 0.5)$ respectively). It can be seen that the approximation gets better as β^* grows both for $q=2$ and $q=3$. For $\beta^* = 2, 3$ the approximation is very good, also in both considered cases of two and three dimensional covariates. Exemplarily we show the results for $n = 20^q$ since the effect is very similar for the other cases.

To evaluate the performance of the estimator on the whole interval $[0,1]^q$ we display in Table 2 the arithmetic mean of the estimated mean squared error of $\hat{g}(\mathbf{x}_1), \ldots, \hat{g}(\mathbf{x}_N)$ where $\mathbf{x}_1, \ldots, \mathbf{x}_N$ form a grid on $[0,1]^q$ with $N = 20^q$. It can be seen that the performance is surprisingly poor for $n = 10^2$ and $\beta^* = 2, 3$. In Figure 4 we show plots of the estimated mean squared error of $\hat{g}$ on $[0,1]^2$ for these cases. From the picture it can be deduced that the problem lies on the boundaries. The reason could be the smaller number of observations in this area which are of higher significance for larger β^*.
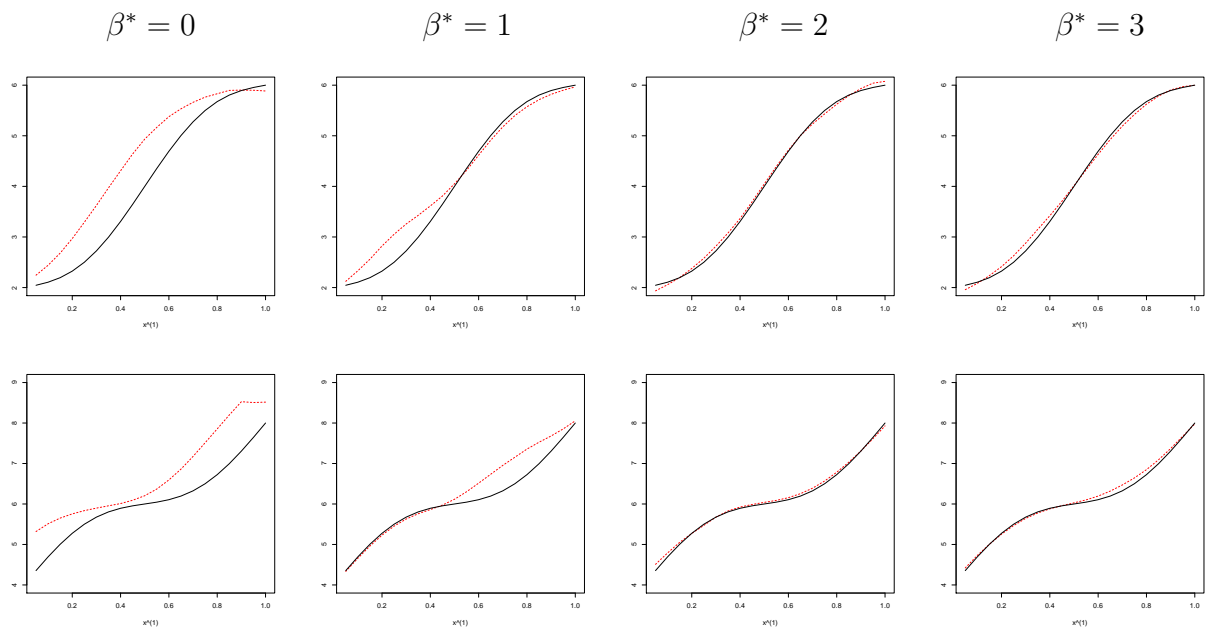
Figure 3: True boundary curve (black solid line) and estimator $\hat{g}$ (red dashed line) for different values of β^*. The plot shows a cut trough the two dimensional surface (top) respectively three dimensional surface (bottom) of the function.
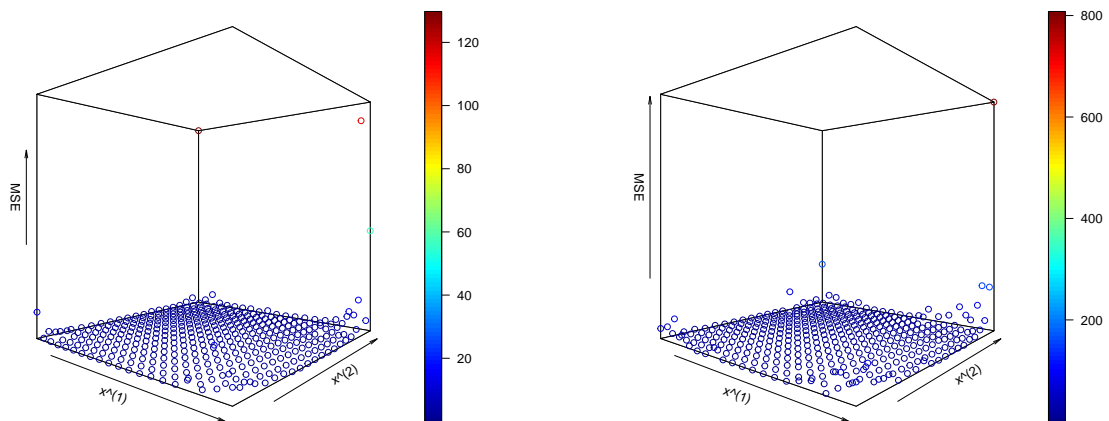

Figure 4: Estimated mean squared error of $\hat{g}$ on $[0,1]^2$ with $\beta^* = 2$ (left-hand side) and $\beta^* = 3$ (right-hand side) and $n = 10^2$.

A   Appendix: Proofs of theorems

A.1   Proof of Theorem 4.1

The proof of Theorem 4.1 is a direct application of Propositions A.1 and A.4 stated below and follows along similar lines as the proof of Theorem 2.2 in Drees et al. (2019).

Proposition A.1 *Assume that model (2.2) holds under (K2) and (K4) and consider the regression function estimator $\hat{g}$ defined in (3.1)-(3.2) where g fulfills condition (K1) for some $\beta \in (0, \beta^* + 1]$ and some $c_g \in [0, c^*]$. Then, there exist $C_{\beta^*, q, c^*}, C_{\beta^*, q}$ and a natural number $J(\beta^*)$, which depend only on the respective subscripts such that for all $\mathbf{x} \in [0, 1]^q$*

$$|\hat{g}(\mathbf{x}) - g(\mathbf{x})| \leq C_{\beta^*, q, c^*} h_n^{\beta} + C_{\beta^*, q} \max_{\mathbf{j} \in \{1, \ldots, 2J(\beta^*)\}^q} \left(\min_{\mathbf{X}_i \in \mathbf{x} + h_n I_{\mathbf{j}}} |\varepsilon_i| \right),$$

where $\mathbf{x} + h_n I_{\mathbf{j}} = \{(x^{(1)} + h_n I_{j_1}) \cap [0, 1]\} \times \ldots \times \{(x^{(q)} + h_n I_{j_q}) \cap [0, 1]\}$ with $I_k = [-1 + (k - 1)/J(\beta^*), -1 + k/J(\beta^*)]$.

Proof: We first highlight that throughout the proof the design points $\mathbf{X}_i : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow ([0, 1]^q, \mathcal{B}[0, 1]^q)$ are random elements. To make the reading easier, we omit the script ω but the proof has to be understood ω-wise that is for any realisation $\mathbf{X}_i(\omega), \omega \in \Omega$. Besides, unless it is specified otherwise, $n \in \mathbb{N}$ is arbitrary.

Let $\mathbf{x} \in [0, 1]^q$ be fixed and for $n \geq 1$ set $I_n^* = \{[x^{(1)} - h_n, x^{(1)} + h_n] \cap [0, 1]\} \times \ldots \times \{[x^{(q)} - h_n, x^{(q)} + h_n] \cap [0, 1]\}$. Note that assumption (K3) here is not required. Nevertheless, we assume that h_n are positive bandwidths such that $\lim_{n \to \infty} h_n = 0$. The idea of proof is based on Theorem 3.1 in Jirak, Meister and Reiß (2014) but comprehensive adaptions are needed to deal with the multivariate case. We consider random design points satisfying assumption (K4) where the Riemann approximation in the aforementioned paper is replaced by the integral defined in (3.2).

This means that we consider the coefficients $(\hat{b}_{\mathbf{j}})_{\mathbf{j}}$ for all multiindices $\mathbf{j}$ with $|\mathbf{j}| \in \{0, \ldots, \beta^*\}$ which minimize the objective function

$$S(\mathbf{x}, (b_{\mathbf{j}})_{\mathbf{j}}) = \int_{I_n^*} \sum_{\mathbf{j} \in \mathbb{N}_0^q : |\mathbf{j}| \leq \beta^*} b_{\mathbf{j}} (\mathbf{t} - \mathbf{x})^{\mathbf{j}} d\mathbf{t} \tag{A.1}$$

under the constraint $Y_i \leq \sum_{\mathbf{j} \in \mathbb{N}_0^q : |\mathbf{j}| \leq \beta^*} b_{\mathbf{j}} (\mathbf{X}_i - \mathbf{x})^{\mathbf{j}}$ for all i with $\mathbf{X}_i \in I_n^*$.

Now, a Taylor-Lagrange development up to the order $\lfloor \beta \rfloor$ of g around $\mathbf{x}$ yields

$$g(\mathbf{X}_i) = \sum_{|\mathbf{j}| \leq \lfloor \beta \rfloor - 1} \frac{D^{\mathbf{j}} g(\mathbf{x})}{\mathbf{j}!} (\mathbf{X}_i - \mathbf{x})^{\mathbf{j}} + \sum_{|\mathbf{j}| = \lfloor \beta \rfloor} \frac{D^{\mathbf{j}} g(\mathbf{x} + \theta(\mathbf{X}_i - \mathbf{x}))}{\mathbf{j}!} (\mathbf{X}_i - \mathbf{x})^{\mathbf{j}}$$

15

with $\theta \in [0, 1]$. Then we have

$$g(\mathbf{X}_i) = \sum_{|\mathbf{j}| \leq \lfloor \beta \rfloor - 1} \frac{D^{\mathbf{j}} g(\mathbf{x})}{\mathbf{j}!} (\mathbf{X}_i - \mathbf{x})^{\mathbf{j}} + \sum_{|\mathbf{j}| = \lfloor \beta \rfloor} \frac{D^{\mathbf{j}} g(\mathbf{x})}{\mathbf{j}!} (\mathbf{X}_i - \mathbf{x})^{\mathbf{j}} + \sum_{|\mathbf{j}| = \lfloor \beta \rfloor} \frac{D^{\mathbf{j}} g(\mathbf{x} + \theta(\mathbf{X}_i - \mathbf{x}))}{\mathbf{j}!} (\mathbf{X}_i - \mathbf{x})^{\mathbf{j}}$$

$$- \sum_{|\mathbf{j}| = \lfloor \beta \rfloor} \frac{D^{\mathbf{j}} g(\mathbf{x})}{\mathbf{j}!} (\mathbf{X}_i - \mathbf{x})^{\mathbf{j}}$$

$$= \sum_{|\mathbf{j}| \leq \lfloor \beta \rfloor} \frac{D^{\mathbf{j}} g(\mathbf{x})}{\mathbf{j}!} (\mathbf{X}_i - \mathbf{x})^{\mathbf{j}} + r_{\lfloor \beta \rfloor}(\mathbf{X}_i, \mathbf{x})$$

where $r_{\lfloor \beta \rfloor}$ is the remainder term defined for $\mathbf{t}, \mathbf{x} \in [0, 1]^q$ by

$$r_{\lfloor \beta \rfloor}(\mathbf{t}, \mathbf{x}) = \sum_{|\mathbf{j}| = \lfloor \beta \rfloor} \frac{D^{\mathbf{j}} g(\mathbf{x} + \theta(\mathbf{t} - \mathbf{x})) - D^{\mathbf{j}} g(\mathbf{x})}{\mathbf{j}!} (\mathbf{t} - \mathbf{x})^{\mathbf{j}}.$$

By assumption (K1) we can make use of the Hölder property (3.4) and get

$$r_{\lfloor \beta \rfloor}(\mathbf{t}, \mathbf{x}) \leq c_g \|\mathbf{t} - \mathbf{x}\|^{\beta - \lfloor \beta \rfloor} \sum_{|\mathbf{j}| = \lfloor \beta \rfloor} \frac{(\mathbf{t} - \mathbf{x})^{\mathbf{j}}}{\mathbf{j}!}$$

$$\leq c_g \|\mathbf{t} - \mathbf{x}\|^{\beta - \lfloor \beta \rfloor} \sum_{|\mathbf{j}| = \lfloor \beta \rfloor} \frac{\|\mathbf{t} - \mathbf{x}\|^{\lfloor \beta \rfloor}}{\mathbf{j}!}$$

$$= c_g c_1(\beta, q) \|\mathbf{t} - \mathbf{x}\|^{\beta}$$

with some constants $c_g, c_1(\beta, q) < \infty$.

Consider now that $b_{\mathbf{j}}$ are the Taylor coefficients such that

$$b_{\mathbf{j}} = \begin{cases} D^{\mathbf{j}} g(\mathbf{x})/\mathbf{j}! & \text{if} \quad |\mathbf{j}| \leq \lfloor \beta \rfloor \\ 0 & \text{if} \quad |\mathbf{j}| > \lfloor \beta \rfloor. \end{cases}$$

With this we define the following two quantities

$$\nabla_i = g(\mathbf{X}_i) - \sum_{|\mathbf{j}| \leq \beta^*} b_{\mathbf{j}} (\mathbf{X}_i - \mathbf{x})^{\mathbf{j}}, \qquad i = 1, \ldots, n$$

and

$$\nabla_n^* := \sup_{\mathbf{t} \in I_n^*} |r_{\lfloor \beta \rfloor}(\mathbf{t}, \mathbf{x})|.$$

Then, one may rewrite the data points Y_i in the model (2.2) as

$$Y_i = \sum_{|\mathbf{j}| \leq \beta^*} b_{\mathbf{j}} (\mathbf{X}_i - \mathbf{x})^{\mathbf{j}} + \varepsilon_i + \nabla_i, \quad i = 1, \ldots, n,$$

16

and from what precedes, under the constraint $\mathbf{X}_i \in I_n^*$, it follows that

$$\nabla_i \leq \nabla_n^* \leq c_g c_1(\beta, q) h_n^\beta. \tag{A.2}$$

Since the errors $\varepsilon_i, i = 1, \ldots, n$ are non-positive, we have for any $\mathbf{X}_i \in I_n^*$

$$
\begin{aligned}
Y_i &= \sum_{|\mathbf{j}| \leq \beta^*} b_{\mathbf{j}} (\mathbf{X}_i - \mathbf{x})^{\mathbf{j}} + \varepsilon_i + \nabla_i \\
&\leq \sum_{|\mathbf{j}| \leq \beta^*} b_{\mathbf{j}} (\mathbf{X}_i - \mathbf{x})^{\mathbf{j}} + \nabla_i \\
&\leq \sum_{|\mathbf{j}| \leq \beta^*} b_{\mathbf{j}} (\mathbf{X}_i - \mathbf{x})^{\mathbf{j}} + \nabla_n^*, \quad i = 1, \ldots, n.
\end{aligned}
$$

Thus, since the coefficients $(\hat{b}_{\mathbf{j}})_{\mathbf{j}}$ minimize the local integral (A.1), we have

$$\int_{I_n^*} \sum_{|\mathbf{j}| \leq \beta^*} \hat{b}_{\mathbf{j}} (\mathbf{t} - \mathbf{x})^{\mathbf{j}} d\mathbf{t} \leq \int_{I_n^*} \left(\sum_{|\mathbf{j}| \leq \beta^*} b_{\mathbf{j}} (\mathbf{t} - \mathbf{x})^{\mathbf{j}} + \nabla_n^* \right) d\mathbf{t} \tag{A.3}$$

for all n. Define now the polynomial

$$Q(\mathbf{t}) = \sum_{|\mathbf{j}| \leq \beta^*} (b_{\mathbf{j}} - \hat{b}_{\mathbf{j}})(\mathbf{t} - \mathbf{x})^{\mathbf{j}} + \nabla_n^*, \quad \mathbf{t} \in I_n^*$$

as the difference of the integrands of the last two quantities. From (A.3), it follows that

$$\int_{I_n^*} Q(\mathbf{t}) d\mathbf{t} \geq 0 \tag{A.4}$$

for any n and for any boundary curve g.

Define now the three sets Q^+, Q^- and Q^0 as

$$Q^+ := \{\mathbf{t} \in I_n^* : Q(\mathbf{t}) > 0\}, \; Q^- := \{\mathbf{t} \in I_n^* : Q(\mathbf{t}) < 0\} \quad \text{and} \quad Q^0 := \{\mathbf{t} \in I_n^* : Q(\mathbf{t}) = 0\}.$$

First note that
$$\lambda(Q^0) = 0 \quad \text{or} \quad Q \equiv 0$$

where λ is the Lebesgue measure. In the latter case, Proposition A.1 is trivially true since then from (A.2) we have

$$|\hat{g}(\mathbf{x}) - g(\mathbf{x})| = |\hat{b}_{\mathbf{0}} - b_{\mathbf{0}}| = \nabla_n^* \leq c_g c_1(\beta, q) h_n^\beta.$$

For $Q \not\equiv 0$ we have

$$\lambda(Q^+) > 0 \tag{A.5}$$

17

which we will prove via a contradiction. If $\lambda(Q^+) = 0$ this would imply that $\lambda(Q^-) = \lambda(I_n^*)$ since $\lambda(Q^0) = 0$ and $Q^- \cup Q^+ \cup Q^0 = I_n^*$. But $\frac{\lambda(Q^-)}{\lambda(I_n^*)} = 1$ is a contradiction to (A.4) so (A.5) must be true. This is needed for the application of Lemma A.2.

Note that by definition

$$Q(\mathbf{x}) = \sum_{|\mathbf{j}|\leq\beta^*}(b_{\mathbf{j}} - \hat{b}_{\mathbf{j}})(\mathbf{0})^{\mathbf{j}} + \nabla_n^* = b_{\mathbf{0}} - \hat{b}_{\mathbf{0}} + \nabla_n^*$$

and thus

$$\sup_{\mathbf{t}\in I_n^*}|Q(\mathbf{t})| \geq |Q(\mathbf{x})| \geq |\hat{b}_{\mathbf{0}} - b_{\mathbf{0}}| - \nabla_n^*. \tag{A.6}$$

Now define

$$\tilde{Q}(\mathbf{t}) = Q\ \Big(\ (1 - t^{(1)})\max((x^{(1)} - h_n), 0) + t^{(1)}\min((x^{(1)} + h_n), 1),$$
$$\ldots, (1 - t^{(q)})\max((x^{(q)} - h_n), 0) + t^{(q)}\min((x^{(q)} + h_n), 1)\Big).$$

It is a Polynomial on $[0,1]^q$ and it inherits from Q the properties of a nonnegative integral $\int_{[0,1]^q}\tilde{Q}(\mathbf{t})d\mathbf{t} \geq 0$ and an area with positive Lebesgue measure where $\tilde{Q}$ is positive. Thus Lemma A.2 can be applied. With this and (A.6) we get

$$\begin{aligned}
c_3(\beta^*, q)(|\hat{b}_{\mathbf{0}} - b_{\mathbf{0}}| - \nabla_n^*) &\leq c_3(\beta^*, q)\sup_{\mathbf{t}\in I_n^*}|Q(\mathbf{t})| \\
&= c_3(\beta^*, q)\sup_{\mathbf{t}\in[0,1]^q}|\tilde{Q}(\mathbf{t})| \\
&\leq \inf_{\mathbf{t}\in B_\gamma}\tilde{Q}(\mathbf{t}) \\
&= \inf_{\mathbf{t}\in B_\delta}Q(\mathbf{t}) \tag{A.7}
\end{aligned}$$

for some constant $c_3(\beta^*, q) > 0$ where $B_\gamma := [\gamma_1^{(1)}, \gamma_2^{(1)}] \times \cdots \times [\gamma_1^{(q)}, \gamma_2^{(q)}]$ and $B_\delta := [\delta_1^{(1)}, \delta_2^{(1)}] \times \cdots \times [\delta_1^{(q)}, \delta_2^{(q)}]$ with $\delta_1^{(r)} = (1 - \gamma_1^{(r)})\max((x^{(r)} - h_n), 0) + \gamma_1^{(r)}\min((x^{(r)} + h_n), 1)$ and $\delta_2^{(r)} = (1 - \gamma_2^{(r)})\max((x^{(r)} - h_n), 0) + \gamma_2^{(r)}\min((x^{(r)} + h_n), 1)$ for some $\gamma_1, \gamma_2 \in [0,1]^q$. Thus B_δ has a volume of minimum size $c_2(\beta^*, q)h_n^q$ for some $c_2(\beta^*, q) > 0$. On the other hand, for any design point $\mathbf{X}_i \in I_n^*$, by definition,

$$\begin{aligned}
Q(\mathbf{X}_i) &= \sum_{|\mathbf{j}|\leq\beta^*}(b_{\mathbf{j}} - \hat{b}_{\mathbf{j}})(\mathbf{X}_i - \mathbf{x})^{\mathbf{j}} + \nabla_n^* \\
&= g(\mathbf{X}_i) - \nabla_i + \nabla_n^* - \sum_{|\mathbf{j}|\leq\beta^*}\hat{b}_{\mathbf{j}}(\mathbf{X}_i - \mathbf{x})^{\mathbf{j}} \\
&\leq g(\mathbf{X}_i) + 2\nabla_n^* - Y_i \\
&= |\varepsilon_i| + 2\nabla_n^* \tag{A.8}
\end{aligned}$$

18

where the last inequality comes from the constraint $Y_i \leq p(\mathbf{X}_i)$ and $|\bigtriangledown_i| \leq \bigtriangledown_n^*$ for all $i = 1, \ldots, n$.

Combining inequalities (A.7) and (A.8), it follows that

$$
\begin{aligned}
c_3(\beta^*, q)|\hat{b}_\mathbf{0} - b_\mathbf{0}| &\leq c_3(\beta^*, q) \bigtriangledown_n^* + \inf_{\mathbf{t} \in B_\delta} Q(\mathbf{t}) \\
&\leq c_3(\beta^*, q) \bigtriangledown_n^* + \min_{\mathbf{X}_i \in B_\delta} Q(\mathbf{X}_i) \\
&\leq c_3(\beta^*, q) \bigtriangledown_n^* + \min_{\mathbf{X}_i \in B_\delta} (|\varepsilon_i| + 2\bigtriangledown_n^*).
\end{aligned} \tag{A.9}
$$

Then, there exists some positive constant $C_{\beta^*, q}$ such that

$$
|\hat{g}(\mathbf{x}) - g(\mathbf{x})| = |\hat{b}_\mathbf{0} - b_\mathbf{0}| \leq C_{\beta^*, q} \bigtriangledown_n^* + C_{\beta^*, q} \min_{\substack{i \in \{1, \ldots, n\} \\ \mathbf{X}_i \in B_\delta}} |\varepsilon_i|.
$$

Using the upper bound of $\bigtriangledown_n^*$ given in (A.2), there exists some positive constant C_{β^*, q, c^*} depending only on β^*, q and c^* such that

$$
|\hat{g}(\mathbf{x}) - g(\mathbf{x})| = |\hat{b}_\mathbf{0} - b_\mathbf{0}| \leq C_{\beta^*, q, c^*} h_n^\beta + C_{\beta^*, q} \min_{\substack{i \in \{1, \ldots, n\} \\ \mathbf{X}_i \in B_\delta}} |\varepsilon_i|.
$$

Choosing a constant $J(\beta^*) \in \mathbb{N}$ depending only on β^* large enough then there exists some $\mathbf{l} = (l_1, \ldots, l_q)$ with $l_r \in \{1, \ldots, 2J(\beta^*)\}$ for $r = 1, \ldots, q$ such that $\mathbf{x} + h_n I_\mathbf{l} := \{(x^{(1)} + h_n I_{l_1}) \cap [0, 1]\} \times \cdots \times \{(x^{(q)} + h_n I_{l_q}) \cap [0, 1]\} \subseteq B_\delta$ where $I_k = [-1 + (k-1)/J(\beta^*), -1 + k/J(\beta^*)], k \geq 0$ and thus

$$
|\hat{g}(\mathbf{x}) - g(\mathbf{x})| \leq C_{\beta^*, q, c^*} h_n^\beta + C_{\beta^*, q} \max_{\mathbf{j} \in \{1, \ldots, 2J(\beta^*)\}^q} \left(\min_{\mathbf{X}_i \in \mathbf{x} + h_n I_\mathbf{j}} |\varepsilon_i| \right),
$$

which is the desired result and concludes the proof.

$\square$

Lemma A.2 *Let β^* and q be natural numbers. There exist positive real numbers δ, c such that for all polynomial functions $P \colon [0, 1]^q \to \mathbb{R}$ of degree β^* with non-negative integral over $[0, 1]^q$ there exists a δ-ball $B_\delta \subseteq [0, 1]^q$ w.r.t. the maximum norm such that $P \geq 0$ on B_δ and*

$$
\inf_{B_\delta} P \geq c \cdot \sup_{[0,1]^q} |P|.
$$

Proof: For all polynomial functions $P \colon [0, 1]^q \to \mathbb{R}$ of degree β^* we have

$$
||\nabla P||_{[0,1]^q, \text{Eucl}} < 4(\beta^*)^2 \cdot \sup_{[0,1]^q} |P|, \tag{A.10}
$$

where $|| \cdot ||_{[0,1]^q,\text{Eucl}}$ denotes the supremum of the Euclidean norm of its argument over the set $[0,1]^q$ and ∇P stands for the gradient of P. This result follows from Theorem 3.1 in Wilhelmsen (1974); see Remark A.3. By the mean value theorem and the Cauchy-Schwarz inequality, this implies that for all $x, y \in [0,1]^q$ we have

$$|P(x) - P(y)| < L \cdot ||x - y|| \sup_{[0,1]^q} |P|, \qquad (A.11)$$

where $|| \cdot ||$ is the maximum norm and $L = 4\sqrt{q}(\beta^*)^2$. The factor $\sqrt{q}$ stems from the change of norms.

Define $\varepsilon = \frac{1}{2L}, \delta = \frac{|B_\varepsilon|}{4L}$ and $c = \frac{|B_\varepsilon|}{4}$. Let P be any such polynomial function. Let P attain its supremum over the set $V_+ = \{x \mid P(x) > 0\}$ at the point x_0, and denote by x_1 the point where P attains its supremum on the whole cube $[0,1]^q$. First, we show

$$|P(x_1)| \leq \frac{2}{|B_\varepsilon|} P(x_0), \qquad (A.12)$$

where $|B_\varepsilon|$ denotes the volume of some ε-ball inside $[0,1]^q$. This step is necessary in case $x_1 \notin V_+$, otherwise it is not needed but the statement is still trivially true. We note that P is non-zero on the ε-ball B_ε centered at x_1. Indeed, taking y to be the nearest point to x_1 such that $P(y) = 0$, Equation (A.11) gives $||x_1 - y|| > \varepsilon$. Next, by the mean value theorem there exists $t_0 \in B_\varepsilon$ such that

$$\int_{B_\varepsilon} |P(t)| \, dt = |P(t_0)||B_\varepsilon|.$$

Applying Equation (A.11) to x_1 and t_0 gives $|P(t_0)| \geq \frac{1}{2}|P(x_1)|$. Hence,

$$\begin{aligned}
|P(x_1)| &\leq 2|P(t_0)| \\
&= \frac{2}{|B_\epsilon|} \int_{B_\epsilon} P(t) \, dt \\
&\leq \frac{2}{|B_\epsilon|} \int_{V_+} P(t) \, dt \\
&\leq \frac{2}{|B_\varepsilon|} P(x_0)|V_+| \\
&\leq \frac{2}{|B_\varepsilon|} P(x_0).
\end{aligned}$$

In the second inequality, we used that the integral of P over $[0,1]^q$ is non-negative. Next, let B_δ denote the δ-ball centered at x_0, and let P attain its infimum over B_δ at the point $x_2 \in B_\delta$. Note that $P > 0$ on B_δ. Indeed, taking y to be the nearest point to x_0 such that $P(y) = 0$, Equations (A.11) and (A.12) give $||x_0 - y|| > \delta$. Finally, applying Equations (A.12) and (A.11) to x_0 and x_2 gives

$$\frac{|B_\varepsilon|}{2}|P(x_1)| - P(x_2) \leq P(x_0) - P(x_2)$$

$$\leq L|P(x_1)|\delta$$
$$= \frac{|B_\varepsilon|}{4}|P(x_1)|.$$

Hence, $P(x_2) \geq c|P(x_1)|$ as required. This concludes the proof.

$\square$

Remark A.3 *The bound given in the Markov inequality in Equation* (A.10) *in Lemma A.2 is far from being optimal. For the sake of completeness, we recall the full statement of Theorem 3.1 in Wilhelmsen (1974). With the above notation, it writes*

$$||\nabla P||_{T,Eucl} \leq \frac{4\beta^{*2}}{\omega(T)} \sup_T |P| \tag{A.13}$$

where T is a compact and convex set with non-empty interior and $\omega(T)$ stands for the thickness of T, that is the minimum distance between two parallel supporting hyperplanes for T. Note that in the context of Lemma A.2, $\omega(T) = \omega([0,1]^q) = 1$. The bound of Equation (A.13) *has been improved later on by Kroó and Révész (1999). They showed that for any convex body (that is a compact convex set with nonempty interior) the constant $\frac{4\beta^{*2}}{\omega(T)}$ may be replaced by $\frac{4\beta^{*2}-2\beta^*}{\omega(T)}$. One may go further observing that the space $[0,1]^q$, under some adequate normalisation, may be seen as a central symmetric convex set (a set is central symmetric if and only if with proper shift it is the unit ball of some norm on $\mathbb{R}^q$) so one may use the result of Sarantopoulos (1991) to get an even smaller bound. We refer to the book of Rassias and Tóth (2014) for an exhaustive treatment of Markov-type inequalities for multivariate polynomials. For the sake of simplicity, we have presented the result with the bound given in Wilhelmsen (1974).*

Proposition A.4 *Assume that model* (2.2) *holds under (K2)-(K4). Then*

$$\sup_{\mathbf{x}\in[0,1]^q} \min_{\substack{i\in\{1,\dots,n\} \\ \mathbf{X}_i\in\mathbf{x}+h_nI}} |\varepsilon_i| = O_\mathbb{P}\left(\left(\frac{|\log(h_n n^{1-q})|}{nh_n^q}\right)^{\frac{1}{\alpha}}\right)$$

for every non-degenerate subinterval $I \subseteq [-1,1]^q$ where we set $\min_\emptyset |\varepsilon_i| := 0$.

Proof: The proof is similar to the proof of Proposition A.6. There are just some preliminaries to consider to deal with the random case.

Note first that under assumption (K4)

$$\lim_{n\to\infty} \frac{|\log(h_n n^{1-q})|}{nh_n^q} = \lim_{n\to\infty} \frac{\log(n)}{nh_n^q}$$

so that Proposition A.4 leads to the appropriate rates in the main result of Theorem 4.1. Similarly Proposition A.6 and Proposition A.7 respectively gives way to the expected rates in the main result of Theorem 4.3.

Now observe that it is obvious that we only have to consider those I where there exists a $d > 0$ such that $|(x^{(r)} + h_n I_r) \cap [0,1]| \geq dh_n$ for all $r = 1, \ldots, q$ with $I = I_1 \times \cdots \times I_q$ and by $|\cdot|$ we mean the length of the one-dimensional intervals.

Next we will show that with probability converging to one there are at least $d_n = O(nh_n^q)$ and at most $d_n' = O(nh_n^q)$ random design points in every hypercube $I_n \subseteq [0,1]^q$ with edge length dh_n. This implies that at least d_n random design points lie in $x + h_n I$. The number of points in I_n can be written as $\sum_{i=1}^n I\{\mathbf{X}_i \in I_n\}$ and for all I_n

$$\frac{1}{n}\sum_{i=1}^n I\{\mathbf{X}_i \in I_n\} \leq \mathbb{P}(\mathbf{X}_1 \in I_n)$$
$$+ \left| \frac{1}{n}\sum_{i=1}^n I\{\mathbf{X}_i \in I_n\} - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n I\{\mathbf{X}_i \in I_n\}\right] \right|,$$
$$\frac{1}{n}\sum_{i=1}^n I\{\mathbf{X}_i \in I_n\} \geq \mathbb{P}(\mathbf{X}_1 \in I_n)$$
$$- \left| \frac{1}{n}\sum_{i=1}^n I\{\mathbf{X}_i \in I_n\} - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n I\{\mathbf{X}_i \in I_n\}\right] \right|.$$

For the term $\mathbb{P}(\mathbf{X}_1 \in I_n)$ we get the upper bound $(dh_n)^q \sup_{\mathbf{t}\in[0,1]^q} f_\mathbf{X}(\mathbf{t})$ and the lower bound $(dh_n)^q \inf_{\mathbf{t}\in[0,1]^q} f_\mathbf{X}(\mathbf{t})$ by

$$\mathbb{P}(\mathbf{X}_1 \in I_n) = \int_{I_n} f_\mathbf{X}(\mathbf{t})d\mathbf{t}$$

and the consideration about the length of I_n. Both bounds on $\mathbb{P}(\mathbf{X}_1 \in I_n)$ are of order $O(h_n^q)$ by assumption (K4). Thus it remains to prove that

$$\sup_{I_n} \left| \frac{1}{n}\sum_{i=1}^n I\{\mathbf{X}_i \in I_n\} - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n I\{\mathbf{X}_i \in I_n\}\right] \right| = O(h_n^q) \tag{A.14}$$

with probability converging to one. Set therefore $P_n f_{n,\mathbf{x}} := \frac{1}{n}\sum_{i=1}^n f_{n,\mathbf{x}}(\mathbf{X}_i)$ and $P f_{n,\mathbf{x}} := \mathbb{E}[f_{n,\mathbf{x}}(\mathbf{X}_1)]$ with $f_{n,\mathbf{x}}(\mathbf{t}) := I\{\mathbf{t} \in I_n\}$. Note that $P f_{n,\mathbf{x}}^2 = \mathbb{P}(\mathbf{X}_1 \in I_n) \leq (dh_n)^q \sup_{\mathbf{t}\in[0,1]^q} f_\mathbf{X}(\mathbf{t})$. Note further that $f_{n,\mathbf{x}}(\mathbf{t}) = I\{\mathbf{t} \in I_n\} = I\{\|\mathbf{t} - \frac{\mathbf{a}_n+\mathbf{b}_n}{2}\| \leq \frac{d}{2}h_n\}$ - where $\mathbf{a}_n = (a_{n,1}, \ldots, a_{n,q})$, $\mathbf{b}_n = (b_{n,1}, \ldots, b_{n,q})$ and $I_n = [a_{n,1}, b_{n,1}] \times \cdots \times [a_{n,q}, b_{n,q}]$ - and thus the conditions of Example 38 and Problem 28 in Pollard (1984) are fullfilled. Then since $|f_{n,\mathbf{x}}| \leq 1$ Theorem 37 in Pollard (1984) can be applied and thus

$$\sup_{I_n} |P_n f_{n,\mathbf{x}} - P f_{n,\mathbf{x}}| = o(h_n^q) \qquad \text{a.\,s.}$$

which proves (A.14). Now the assertion of the Proposition follows with the same arguments as in the proof of Proposition A.6 with the modification from the proof of Proposition A.7.

$\square$

A.2   Proof of Theorem 4.3

The proof of Theorem 4.3 is similar to the proof of Theorem 4.1 and is based on the following Propositions A.5, A.6 and A.7 respectively.

Proposition A.5 *Assume that model* (2.3) *holds under (K2) and (K4') and let $\hat{g}$ be defined in* (3.5)-(3.6) *satisfying (K1) for some $\beta \in (0, \beta^* + 1]$ and some $c_g \in [0, c^*]$. Then, there exists constants $C_{\beta^*,q,c^*}, C_{\beta^*,q}$ and a natural number $J(\beta^*)$ such that for all $\mathbf{x} \in [0,1]^q$*

$$|\hat{g}(\mathbf{x}) - g(\mathbf{x})| \leq C_{\beta^*,q,c^*} h_n^{\beta} + C_{\beta^*,q} \max_{\mathbf{j} \in \{1,\ldots,2J(\beta^*)\}^q} \left(\min_{\substack{i \in \{1,\ldots,n\} \\ \mathbf{x}_{i,n} \in \mathbf{x} + h_n I_{\mathbf{j}}}} |\varepsilon_{i,n}| \right),$$

where $\mathbf{x} + h_n I_{\mathbf{j}} = \{(x^{(1)} + h_n I_{j_1}) \cap [0,1]\} \times \cdots \times \{(x^{(q)} + h_n I_{j_q}) \cap [0,1]\}$ with $I_k = [-1 + (k - 1)/J(\beta^*), -1 + k/J(\beta^*)]$.

Proof: The proof is almost identical to the proof of Proposition A.1 and is skipped here for the sake of conciseness.

Proposition A.6 *Assume that model* (2.3) *holds with equidistant design points under (K2) and (K3') (which is equivalent to (K3) in this case). Then*

$$\sup_{\mathbf{x} \in [0,1]^q} \min_{\substack{i \in \{1,\ldots,n\} \\ \mathbf{x}_{i,n} \in \mathbf{x} + h_n I}} |\varepsilon_{i,n}| = O_{\mathbb{P}}\left(\left(\frac{\left| \log\left(h_n n^{\frac{1-q}{q}} \right) \right|}{n h_n^q} \right)^{\frac{1}{\alpha}} \right)$$

for every non-degenerate subinterval $I \subseteq [-1,1]^q$ where we set $\min_{\emptyset} |\varepsilon_{i,n}| := 0$.

Proof: The proof is similar to the proof of Lemma A.2 in Drees et al. (2019) and the proof of Lemma A.1 in Neumeyer et al. (2019) but comprehensive adaptions are needed to deal with the multivariate case.

Let $Z_1, Z_2, \ldots$ be iid with the same distribution as $-\varepsilon_{i,n}$ with cumulative distribution function U. Recall that the distribution of the errors $\varepsilon_{i,n}$ do not depend of n, that is why, for the sake of clarity, the second index n is omitted in the Z_i. To prove the result we shall show that

$$\exists L < \infty : \quad \mathbb{P}\left(\sup_{\mathbf{x} \in [0,1]^q} \min_{\substack{i \in \{1,\ldots,n\} \\ \mathbf{x}_{i,n} \in \mathbf{x} + h_n I}} Z_i > L r_n \right) \xrightarrow[n \to \infty]{} 0$$

with $r_n = \left(\frac{\left| \log\left(h_n n^{\frac{1-q}{q}} \right) \right|}{n h_n^q} \right)^{\frac{1}{\alpha}}$.
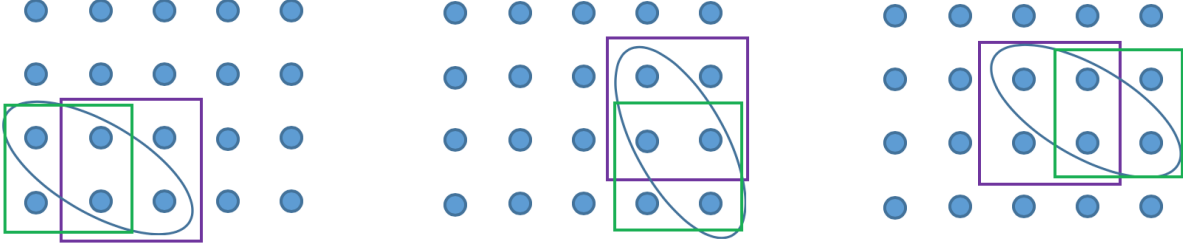
Figure 5: Example for building sets for $q = 2, d_n = 4$. On the left hand side the first three sets are illustrated; in the middle and on the right hand side the transition from one row to the next one is shown. In each picture the smaller green rectangle indicates $D_{j-1,n}$, the oval $D_{j,n}$ and the larger purple rectangle $D_{j+1,n}$, respectively.

It is obvious that we only have to consider those I where there exists a $d > 0$ such that $|(x^{(r)} + h_n I_r) \cap [0,1]| \geq dh_n$ for all $r = 1, \ldots, q$ with $I = I_1 \times \cdots \times I_q$ and by $|\cdot|$ we mean the length of the one-dimensional intervals. Note that at least $d_n = \lfloor dh_n n^{\frac{1}{q}} \rfloor^q$ design points lie in such an I.

Now we want to arrange the design points $\mathbf{x}_{i,n}$ in sets of size d_n such that we can replace $\sup_{\mathbf{x} \in [0,1]^q}$ by a maximum over these sets. To this end set

$$D_{1,n} := \{\mathbf{x}_{i,n} : x_{i,n}^{(r)} \leq dh_n, r = 1, \ldots, q\}.$$

For $j = 2, 3, \ldots$ the set $D_{j,n}$ is built out of $D_{j-1,n}$ by removing one design point from the set and adding a new one in an appropriate way, so that all sets $D_{j,n}$ fulfill:

(E1) For every $\mathbf{x} \in [0,1]^q$ there exists $D_{j,n}$ with $D_{j,n} \subseteq \{\mathbf{x}_{i,n} : \mathbf{x}_{i,n} \in \mathbf{x} + h_n I\}$.

(E2) Each set $D_{j,n}$ contains exactly d_n design points.

(E3) For every $j \in \{2, \ldots, \widetilde{d}_n\}$ there exists exactly one i with $\mathbf{x}_{i,n} \in D_{j-1,n} \backslash D_{j,n}$, where $\widetilde{d}_n$ is the total number of sets.

An example for building these sets is given in Figure 5. With this procedure

$$\widetilde{d}_n = \left((n^{\frac{1}{q}} - d_n^{\frac{1}{q}} + 1)^q - 1\right)d_n^{\frac{q-1}{q}} + 1$$

since there are $(n^{\frac{1}{q}} - d_n^{\frac{1}{q}} + 1)^q$ sets in rectangles and $d_n^{\frac{q-1}{q}}$ steps are needed to get from one rectangle to the next one.

Now the supremum can be replaced by maximum, i. e. for all $y > 0$

$$\mathbb{P}\left(\sup_{\substack{\mathbf{x} \in [0,1]^q}} \min_{\substack{i \in \{1, \ldots, n\} \\ \mathbf{x}_{i,n} \in \mathbf{x} + h_n I}} Z_i > y\right) \leq \mathbb{P}\left(\max_{\substack{j \in \{1, \ldots, \widetilde{d}_n\}}} \min_{\substack{i \in \{1, \ldots, n\} \\ \mathbf{x}_{i,n} \in D_{j,n}}} Z_i > y\right)$$

24

$$= \mathbb{P}\left(\max_{l \in \{0,\dots,l_n\}} M_{l,n} > y \right)$$

where $l_n = \langle \frac{\widetilde{d_n}}{d_n} \rangle$ and

$$M_{l,n} := \max_{\substack{j \in \{ld_n+1,\dots,(l+1)d_n\}}} \min_{\substack{i \in \{1,\dots,n\} \\ \mathbf{x}_{i,n} \in D_{j,n}}} Z_i \qquad \text{for } l = 0,\dots,l_n-1$$

as well as $M_{l_n,n} := \max_{j \in \{l_n d_n+1,\dots,\widetilde{d_n}\}} \min_{\substack{i \in \{1,\dots,n\} \\ \mathbf{x}_{i,n} \in D_{j,n}}} Z_i$.

The random variables $M_{l,n}$ are then sectioned in k_n groups such that those $M_{l,n}$ that are in one group contain different Z_i and therefore are independent. This is possible with $k_n = o(l_n)$ groups and

$$\mathbb{P}\left(\sup_{\mathbf{x} \in [0,1]^q} \min_{\substack{i \in \{1,\dots,n\} \\ |\mathbf{x}_{i,n}-\mathbf{x}| \le h_n}} Z_i > y \right) \le \sum_{j=1}^{k_n} \mathbb{P}\left(\max_{\substack{l \in \{0,\dots,l_n\} \\ l \text{ in group } j}} M_{l,n} > y \right)$$

$$\le k_n \left(1 - \mathbb{P}(M_{0,n} \le y)^{\lceil \frac{l_n+1}{k_n} \rceil} \right).$$

Note that for every $j \in \{2,\dots,d_n\}$ there exists exactly one i with $\mathbf{x}_{i,n} \in D_{j-1,n} \backslash D_{j,n}$. We denote this by $i(j)$. Now analogous to the proof of Lemma A.1 in Neumeyer et al. (2019) we have

$$\mathbb{P}(M_{0,n} > y) = \mathbb{P}\left(\min_{\substack{i \in \{1,\dots,n\} \\ \mathbf{x}_{i,n} \in D_{1,n}}} Z_i > y \right) + \sum_{j=2}^{d_n} \mathbb{P}\left(\{Z_{i(j)} \le y\} \cap \left\{ \min_{\substack{i \in \{1,\dots,n\} \\ \mathbf{x}_{i,n} \in D_{j,n}}} Z_i > y \right\} \right)$$

$$= \overline{U}(y)^{d_n} + (d_n - 1)U(y)\overline{U}(y)^{d_n}$$

$$\le (1 + d_n U(y))\overline{U}(y)^{d_n}$$

and it remains to show that for sufficiently large L

$$k_n \left(1 - \left(1 - (1 + d_n U(Lr_n))\overline{U}(Lr_n)^{d_n} \right)^{\lceil \frac{l_n+1}{k_n} \rceil} \right) \xrightarrow[n \to \infty]{} 0 \tag{A.15}$$

which is true if

$$(1 + d_n U(Lr_n))\overline{U}(Lr_n)^{d_n} = o\left(\frac{d_n}{\widetilde{d_n}} \right). \tag{A.16}$$

This implication follows from several Taylor expansions. Let $z_n \to 0$ and $y_n \to \infty$ denote some sequences with $y_n z_n \to 0$. Now by a second order Taylor expansion of $\log(1 + z_n)$ we get for n sufficiently large

$$k_n(1 - (1 + z_n)^{y_n}) = k_n(1 - \exp(\log((1 + z_n)^{y_n})))$$

$$= k_n(1 - \exp(y_n z_n) + O(y_n z_n^2))$$

$$= k_n(1 - \exp(0) + O(y_n z_n) + O(y_n z_n^2))$$

25

$$= O(k_n y_n z_n)$$

where the third equality follows by a first order Taylor expansion of $\exp(y_n z_n)$. This means that

$$k_n \left(1 - \left(1 - (1 + d_n U(Lr_n))\overline{U}(Lr_n)^{d_n} \right)^{\lceil \frac{l_n+1}{k_n} \rceil} \right) = O\left(k_n \frac{l_n}{k_n} (1 + d_n U(Lr_n))\overline{U}(Lr_n)^{d_n} \right)$$

which proves our claim $(A.16) \Rightarrow (A.15)$.

Note that $U(Lr_n) = aL^\alpha \frac{\left| \log\left(n^{-1}(nh_n^q)^{1/q} \right) \right|}{nh_n^q}(1 + o(1))$ for some positive constant a and thus we have

$$\begin{aligned}
\overline{U}(Lr_n)^{d_n} &= \exp(d_n \log(1 - U(Lr_n))) \\
&\underset{n\to\infty}{\sim} \exp\left(-aL^\alpha \left| \log\left(n^{-1}(nh_n^q)^{1/q} \right) \right| \right)
\end{aligned}$$

by a second order Taylor expansion of $\log(1 - U(Lr_n))$ and $d_n \underset{n\to\infty}{\sim} nh_n^q$. Further

$$\begin{aligned}
(1 + d_n U(Lr_n))\overline{U}(Lr_n)^{d_n} &\underset{n\to\infty}{\sim} \exp\left(\log\left(\left| \log\left(n^{-1}(nh_n^q)^{1/q} \right) \right| \right) - aL^\alpha \left| \log\left(n^{-1}(nh_n^q)^{1/q} \right) \right| \right) \\
&= \exp\left(\log\left(\left| \log\left(n^{-1}(nh_n^q)^{1/q} \right) \right| (n^{-1}(nh_n^q)^{1/q})^{aL^\alpha} \right) \right) \\
&= h_n n^{\frac{1-q}{q}} \left| \log\left(n^{-1}(nh_n^q)^{1/q} \right) \right| (n^{-1}(nh_n^q)^{1/q})^{aL^\alpha - 1} \\
&= o(h_n n^{\frac{1-q}{q}})
\end{aligned}$$

for sufficiently large L. This concludes the proof since for n sufficiently large

$$\begin{aligned}
h_n n^{\frac{1-q}{q}} &\underset{n\to\infty}{\sim} \frac{d_n^{\frac{1}{q}}}{n} \leq \frac{d_n^{\frac{1}{q}}}{\left(n^{\frac{1}{q}} - d_n^{\frac{1}{q}} \right)^q - 1} \\
&= \frac{d_n}{\left(\left(n^{\frac{1}{q}} - d_n^{\frac{1}{q}} \right)^q - 1 \right) d_n^{\frac{q-1}{q}}} \\
&= O\left(\underbrace{\frac{d_n}{\left((n^{\frac{1}{q}} - d_n^{\frac{1}{q}} + 1)^q - 1 \right) d_n^{\frac{q-1}{q}} + 1}}_{= \widetilde{d_n}} \right).
\end{aligned}$$

$\square$

Proposition A.7 *Assume that model (2.3) holds under (K2), (K3') and (K4'). Then*

$$\sup_{\mathbf{x}\in[0,1]^q} \min_{\substack{i\in\{1,\dots,n\} \\ \mathbf{x}_{i,n}\in\mathbf{x}+h_n I}} |\varepsilon_{i,n}| = O_{\mathbb{P}}\left(\left(\frac{\left| \log\left(d_n d_n'^{-q} h_n^{q^2-q+1} \right) \right|}{d_n} \right)^{\frac{1}{\alpha}} \right)$$

with $d_n = d_n(1)$, $d'_n = d'_n(1)$, for every non-degenerate subinterval $I \subseteq [-1,1]^q$ where we set $\min_\emptyset |\varepsilon_{i,n}| := 0$. If $d_n = nh_n^q$ then $d_n \approx d'_n$ and the convergence rate simplifies to

$$\left(\frac{\left| \log\left(n^{1-q} h_n \right) \right|}{nh_n^q} \right)^{\frac{1}{\alpha}}.$$

Proof: As in the proof of Proposition A.6 it is obvious that we only have to consider those I where there exists a $d > 0$ such that $|(x^{(r)} + h_n I_r) \cap [0,1]| \geq dh_n$ for all $r = 1, \ldots, q$. By assumption (K4') there lie at least $d_n(d)$ design points in such an $\mathbf{x} + h_n I$. In the remainder of the proof we will suppress the dependence on d of d_n and d'_n since it has no influence on the rates.

The proof is similar to the proof of Proposition A.6 but the technique for building the sets $D_{j,n}$ has to be adjusted. The sets shall fulfill the conditions (E1)-(E3) but since the design points are not necessarily equidistant the total number of sets needed is higher. The procedure for building the sets $D_{j,n}$ is based on the same idea as the one shown in Figure 5 for the equidistant case. Since there are no "rows" in the non-equidistant case we pass through the "stripes" $[0,1] \times [0, dh_n]$, $[0,1] \times (\min_i x_{i,n}^{(2)}, \min_i x_{i,n}^{(2)} + dh_n], \ldots$ instead. With this procedure the total number of sets $\widetilde{d}_n$ can be bounded by

$$\widetilde{d}_n \leq \left(\left(\left\lceil \frac{1}{dh_n} \right\rceil d'_n - d_n + 1 \right) + d_n \right) \cdot \prod_{r=2}^{q} \left(\left\lceil \frac{1}{dh_n} \right\rceil^r d'_n - \left\lceil \frac{1}{dh_n} \right\rceil^{r-1} d_n + 1 \right)$$

$$\leq \left(\left\lceil \frac{1}{dh_n} \right\rceil d'_n + 1 \right) \left(\left\lceil \frac{1}{dh_n} \right\rceil^{q-1} \left(\left\lceil \frac{1}{dh_n} \right\rceil d'_n - d_n \right) + 1 \right)^{q-1}$$

for n sufficiently large, since $\left(\left\lceil \frac{1}{dh_n} \right\rceil d'_n - d_n + 1 \right)$ is the maximal number of sets in one "stripe" and $\prod_{r=2}^{q} \left(\left\lceil \frac{1}{dh_n} \right\rceil^r d'_n - \left\lceil \frac{1}{dh_n} \right\rceil^{r-1} d_n + 1 \right)$ is the maximal number of "stripes". Further

$$d_n d_n'^{-q} h_n^{q^2-q+1} = \frac{d_n h_n^{(q-1)^2+1}}{d'_n (h_n^{-1} d'_n)^{q-1}}$$

$$\leq \frac{d_n h_n^{(q-1)^2+1}}{d'_n (h_n^{-1} d'_n - d_n)^{q-1}}$$

$$= \frac{d_n}{h_n^{-1} d'_n h_n^{-(q-1)^2} (h_n^{-1} d'_n - d_n)^{q-1}}$$

$$= O\left(\frac{d_n}{\left(\left\lceil \frac{1}{dh_n} \right\rceil d'_n + 1 \right) \left(\left\lceil \frac{1}{dh_n} \right\rceil^{q-1} \left(\left\lceil \frac{1}{dh_n} \right\rceil d'_n - d_n \right) + 1 \right)^{q-1}} \right)$$

$$= O\left(\frac{d_n}{\widetilde{d}_n} \right).$$

27

Now the assertion follows with the same arguments as in the proof of Proposition A.6.

$\square$

Acknowledgements

References

Aigner, D., Lovell, C. K. and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of econometrics*, **6**(1), 21-37.

Andersen, E. D. and Andersen, K. D. (2000). The MOSEK interior point optimizer for linear programming: an implementation of the homogeneous algorithm. *High Performance Optimization*, 197–232

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B*, **26**, 211–252.

Brown, L.D. and Low M.G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *The annals of statistics*, **24**, 2384–2398.

Cooper, W., Seiford, L. M. and Zhu, J. (Eds.) (2011). *Handbook on data envelopment analysis* (Vol. 164). Springer Science and Business Media.

Cornwell, C. and Schmidt, P. (2008). Stochastic frontier analysis and efficiency estimation. *In The Econometrics of Panel Data* 697-726. Springer, Berlin, Heidelberg.

Daouia, A. and Simar, L. (2005). Robust nonparametric estimators of monotone boundaries. *Journal of Multivariate Analysis*, **96**(2), 311-331.

Daouia, A., Noh, H. and Park, B. U. (2016). Data envelope fitting with constrained polynomial splines. *Journal of the Royal Statistical Society, Series B* **78**, 3–30.

Deprins, D., Simar, L. and Tulkens, H. (1984). *Measuring labor-efficiency in post offices*. No 571, CORE Discussion Papers RP, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).

de Haan, L. and Resnick, S. (1994). Estimating the home range. *Journal of Applied Probability*, **31**, 700–720.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An introduction.* Springer Series in Operations Research and Financial Engineering. Springer, New York.

Drees, H., Neumeyer, N. and Selk, L. (2019). Estimation and hypotheses tests in boundary regression models. *Bernoulli*, **25**, 424–463.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society: Series A (General)*, **120**(3), 253-281.

Francis, J. C. and Kim, D. (2013). *Modern portfolio theory: Foundations, analysis, and new developments* (Vol. 795). John Wiley & Sons.

Gardes, L. (2002). Estimating the support of a poisson process via the faber-schauder basis and extreme values. In *Annales de l'ISUP* **46**(1-2), 43–72.

Gijbels, I., Mammen, E., Park, B.U. and Simar, L. (1999). On estimation of monotone and concave frontier functions. *Journal of American Statistical Association*, **94**, 220–228.

Gijbels, I. and Peng, L. (2000). Estimation of a support curve via order statistics. *Extremes*, **3**, 251–277.

Girard, S. and Jacob, P. (2003). Projection estimates of point processes boundaries. *Journal of Statistical Planning and Inference*, **116**(1), 1–15.

Girard, S. and Jacob, P. (2004). Extreme values and kernel estimates of point processes boundaries. *ESAIM: Probability and Statistics*, **8**, 150–168.

Girard, S., Iouditski, A. and Nazin, A. V. (2005). L1-optimal nonparametric frontier estimation via linear programming. *Automation and Remote Control*, **66**(12), 2000–2018.

Girard, S. and Jacob, P. (2008). Frontier estimation via kernel regression on high power-transformed data. *Journal of Multivariate Analysis* **99**, 403–420.

Girard, S., Guillou, A. and Stupfler, G. (2013). Frontier estimation with kernel regression on high order moments. *Journal of Multivariate Analysis*, **116**, 172–189.

Goetzmann, W. N., Brown, S. J., Gruber, M. J. and Elton, E. J. (2014). *Modern portfolio theory and investment analysis.* John Wiley & Sons, 237.

Hall, P., Nussbaum, M. and Stern, S.E. (1997). On the estimation of a support curve of indeterminate sharpness. *Journal of Multivariate Analysis*, **62**, 204–232.

Hall, P., Park, B.U. and Stern, S.E. (1998). On polynomial estimators of frontiers and boundaries. *Journal of Multivariate Analysis*, **66**, 71–98.

Hall, P. and Park, B.U. (2004). Bandwidth choice for local polynomial estimation of smooth boundaries. *Journal of Multivariate Analysis*, **91** (2), 240—261.

Hall, P. and Van Keilegom, I. (2009). Nonparametric "regression" when errors are positioned at end-points. *Bernoulli*, **15**, 614–633.

Härdle, W., Park, B.U. and Tsybakov, A.B. (1995). Estimation of non-sharp support boundaries. *Journal of Multivariate Analysis*, **55**, 205—218.

Jacob, P. and Suquet, C. (1995). Estimating the edge of a Poisson process by orthogonal series. *Journal of Statistical Planning and Inference*, **46**(2), 215–234.

Jirak, M., Meister, A. and Reiß, M. (2014). Adaptive estimation in nonparametric regression with one-sided errors. *Journal of approximation theory*, **11**, 216–220.

Jones, M.C. and Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika*, **96**, 761–780.

Knight, K. (2001). Limiting distributions of linear programming estimators. *Extremes*, **4** (2), 87–103.

Korostelëv A.P. and Tsybakov A.B. (1993). *Minimax Lower Bounds. In: Minimax Theory of Image Reconstruction*. Lecture Notes in Statistics, **82**. Springer, New York, NY

Korostelëv, A. P., Simar, L. and Tsybakov, A. B. (1995). Efficient estimation of monotone boundaries. *The Annals of Statistics*, **23**(2), 476-489.

Kroó, A. and Révész, S. (1999). On Bernstein and Markov-type inequalities for multivariate polynomials on convex bodies. *Journal of approximation theory*, **99**(1), 134-152.

Kumbhakar, S. C. and Lovell, C. K. (2003). *Stochastic frontier analysis*. Cambridge university press.

Kumbhakar, S. C., Park, B. U., Simar, L. and Tsionas, E. G. (2007). Nonparametric stochastic frontiers: a local maximum likelihood approach. *Journal of Econometrics*, **137**(1), 1–27.

Linton, O., Sperlich, S. and Van Keilegom, I. (2008). Estimation on a semiparametric transformation model. *Annals of Statisics*, **36**, 686–718.

Lovell, C. K., Grosskopf, S., Ley, E., Pastor, J. T., Prior, D. and Eeckaut, P. V. (1994). Linear programming approaches to the measurement and analysis of productive efficiency. *Top*, **2**(2), 175–248.

Mardani, A., Streimikiene, D., Balezentis, T., Saman, M. Z. M., Nor, K. M. and Khoshnava, S. M. (2018). Data envelopment analysis in energy and environmental economics: An overview of the state-of-the-art and recent development trends. *Energies* 2018, **11**(8).

Meister, A. and Reiß, M. (2013). Asymptotic equivalence for nonparametric regression with non-regular errors. *Probability theory and relative fields*, **155**, 201–229.

Menneteau, L. (2008). Multidimensional limit theorems for smoothed extreme value estimates of point processes boundaries. *ESAIM: Probability and Statistics*, **12**, 273-307.

Meeusen, W. and van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International economic review*, 435-444.

Müller, U.U. and Wefelmeyer W. (2010). Estimation in nonparametric regression with non-regular errors. *Communications in Statistics - Theory and Methods*, **39**, 1619–1629.

Neumeyer, N., Selk, L. and Tillier, C. (2019). Semi-parametric transformation boundary regression models. *Annals of the Institute of Statistical Mathematics*, **27**, 1–29.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

Ramanathan, R. (2003). *An introduction to data envelopment analysis: a tool for performance measurement.* Sage.

Rassias, T. M. and Tóth, L. (Eds.). (2014). *Topics in Mathematical Analysis and Applications.* Springer International Publishing.

Reiß, M. and Selk, L. (2017). Efficient estimation of functionals in nonparametric boundary models. *Bernoulli*, **23**(2), 1022–1055.

Sarantopoulos, Y. (1991). Bounds on the derivatives of polynomials on Banach spaces. *Mathematical Proceedings of the Cambridge Philosophical Society*, **110**(2), 307–312.

Simar, L. and Wilson, P.W. (1998). Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Management Science*, **44**, 49–61.

Simar, L. and Zelenyuk, V. (2011). Stochastic FDH/DEA estimators for frontier analysis. *Journal of Productivity Analysis*, **36**(1), 1–20.

Vaidya, P. M. (1989). Speeding-up linear programming using fast matrix multiplication. *30th Annual Symposium on Foundations of Computer Science*, 332–337.

Wilhelmsen, D. R. (1974). A Markov inequality in several dimensions. *Journal of Approximation Theory*, **11**(3), 216–220.

Wilson, P.W. (2003). Testing independence in models of productive efficiency. *Journal of Productivity Analysis*, **20**, 361–390.