

# Semi-parametric transformation boundary regression models\*

NATALIE NEUMEYER, LEONIE SELK and CHARLES TILLIER

Department of Mathematics, University of Hamburg

January 31, 2019

## Abstract

In the context of nonparametric regression models with one-sided errors, we consider parametric transformations of the response variable in order to obtain independence between the errors and the covariates. In view of estimating the transformation parameter, we use a minimum distance approach and show the uniform consistency of the estimator under mild conditions. The boundary curve, i.e. the regression function, is estimated applying a smoothed version of a local constant approximation for which we also prove the uniform consistency. We deal with both cases of random covariates and deterministic (fixed) design points. To highlight the applicability of the procedures and to demonstrate their performance, the small sample behavior is investigated in a simulation study using the so-called Yeo-Johnson transformations.

**Key words:** Box-Cox transformations, frontier estimation, minimum-distance estimation, local constant approximation, boundary models, nonparametric regression, Yeo-Johnson transformations

## 1 Introduction

Before fitting a regression model it is very common in applications to transform the response variable. The aim of the transformation is to gain efficiency in the statistical inference, for instance, by reducing skewness or inducing a specific structure of the model, e.g. linearity

---

\*Financial support by the DFG (Research Unit FOR 1735 *Structural Inference in Statistics: Adaptation and Efficiency*) is gratefully acknowledged.

of the regression function or homoscedasticity. In practice often a parametric class of transformations is considered from which an ‘optimal’ one should be selected data-dependently (with a specific purpose in mind). A classical example is the class of Box-Cox power transformations introduced for linear models by Box and Cox (1964). There is a vast literature on parametric transformation models in the context of mean regression and we refer to the monograph by Carroll and Ruppert (1988). Powell (1991) introduced Box-Cox transformations in the context of linear quantile regression; see also Mu and He (2007) who considered transformations to obtain a linear quantile regression function. Horowitz (2009) reviewed estimation in transformation models with parametric regression in the cases where either the transformation or the error distribution or both are modeled nonparametrically. Linton, Sperlich and Van Keilegom (2008) suggested parametric estimators for transformations, while the error distribution is estimated nonparametrically and the regression function is additive. In this paper, the aim of the transformation is to induce independence between the covariables and the errors. Linton et al. (2008) considered profile likelihood and minimum distance estimation for the transformation parameter. The results for the profile likelihood estimator were generalized for nonparametric regression models by Colling and Van Keilegom (2016).

All literature cited above is about mean or quantile regression. In contrast in the paper at hand we consider boundary regression models. Such nonparametric regression models with one-sided errors have been considered, among others, by Hall and Van Keilegom (2009), Meister and Reiß (2013), Jirak, Meister and Reiß (2014) and Drees, Neumeyer and Selk (2018). Relatedly, estimation of support boundaries have been considered, for instance, by Härdle, Park and Tsybakov (1995), Hall, Park and Stern (1998), Girard and Jacob (2008) and Daouia, Noh and Park (2016). Such models naturally appear when analyzing auctions or records or production frontiers. Unlike conditional mean models, regression models with one-sided errors (as well as quantile regression models) have the attractive feature of equivariance under monotone transformations. Thus in such a model with monotone transformation of the response one can recover the original functional dependence in an easy manner. Similar to Linton et al. (2008) the aim of our transformation is to induce a model where the error distribution does not depend on the covariates. Independence of errors and covariates is a very typical assumption in regression models. For boundary models this assumption is met, e.g., by Müller and Wefelmeyer (2010), Meister and Reiß (2013), and Drees et al. (2018). A transformation inducing (approximate) independence between the covariable and the error would allow for a global bandwidth selection in the adaptive regression estimator suggested by Jirak et al. (2014). Wilson (2003) pointed out that in production frontier models, independence assumptions are needed for validity of bootstrap procedures for nonparametric frontier models (see Simar and Wilson, 1998) and suggested some tests for independence of

errors and covariates (see also Drees et al., 2018).

While Linton et al. (2008) found advantages of the profile likelihood approach over minimum distance estimation of the transformation parameter in corresponding mean regression transformation models, this is at the cost of strong regularity conditions, among others a bounded error density with bounded derivative. In the context of boundary models with error distribution which is regularly varying at zero and irregular, one needs to avoid assumptions on bounded densities. Thus we investigate a minimum distance approach to estimate the transformation parameter and give mild model assumptions under which the estimator is consistent.

We consider the cases of random covariates and deterministic (fixed) design points, which are both meaningful. The equidistant fixed design - as well as its natural generalization to deterministic covariates - is often used in real-life applications when time is involved in the data set. This is the case for instance in Jirak et al. (2014) where the authors studied the monthly sunspot observations and the annual best running times of 1500 meters. 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 insurance and financial risk modelling when analyzing the optimality of portfolios (see Markowitz (1952) for the seminal contribution).

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 consistency of the transformation parameter estimator. In section 5 we present simulation results. The proofs for the random covariate case are given in the appendix, while supplementary material contains proofs for the fixed design case and some additional figures and simulation results.

## 2 Model

### 2.1 The random design case

Consider independent and identically distributed observations  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ , with the same distribution as  $(X, Y)$ , where  $Y$  is univariate and  $X$  is distributed on  $[0, 1]$ . Further consider a family  $\mathcal{L} = \{\Lambda_\vartheta | \vartheta \in \Theta\}$  of strictly increasing and continuous transformations. Throughout the paper we assume existence of a transformation  $\Lambda_{\vartheta_0}$  in the class  $\mathcal{L}$  such that in the corresponding boundary regression model

$$\Lambda_{\vartheta_0}(Y) = h_{\vartheta_0}(X) + \varepsilon \tag{2.1}$$

the errors and the covariates are stochastically independent. Note that for notational simplicity we set  $\Lambda_0 = \Lambda_{\vartheta_0}$  and  $h_0 = h_{\vartheta_0}$ . Further denote by  $F_0$  the cumulative distribution function (cdf) of the independent and identically distributed (iid)  $\varepsilon_i = \Lambda_0(Y_i) - h_0(X_i)$ ,  $i = 1, \dots, n$ . Then we assume that  $F_0(0) = 1$  and  $F_0(-\Delta) < 1$  for all  $\Delta > 0$ . This identifies the function  $h_0$  as the upper boundary curve of the observations since

$$\begin{aligned} \mathbb{P}(\Lambda_0(Y_i) \leq h_0(X_i) \mid X_i = x) &= 1 \text{ for all } x \in [0, 1] \\ \mathbb{P}(\Lambda_0(Y_i) - h_0(X_i) \leq -\Delta \mid X_i = x) &< 1 \text{ for all } x \in [0, 1], \Delta > 0. \end{aligned}$$

The aim is to estimate  $\vartheta_0$  from the observations.

**Remark 2.1** *Note that even if the model does not hold exactly (i.e. there does not exist any  $\vartheta_0 \in \Theta$  that leads to exact independence of the errors and covariates) the transformation can be useful in applications because it will reduce the dependence.*

For each  $\vartheta \in \Theta$  one can consider the transformed responses  $\Lambda_\vartheta(Y_i)$ . Note that those form a boundary regression model with boundary curve  $h_\vartheta = \Lambda_\vartheta \circ \Lambda_0^{-1} \circ h_0$ , because

$$\mathbb{P}(\Lambda_\vartheta(Y_i) \leq h_\vartheta(X_i) \mid X_i = x) = \mathbb{P}(\Lambda_0(Y_i) \leq h_0(X_i) \mid X_i = x) = 1$$

and for each  $\delta > 0$ ,

$$\mathbb{P}(\Lambda_\vartheta(Y_i) - h_\vartheta(X_i) \leq -\delta \mid X_i = x) = \mathbb{P}(\Lambda_0(Y_i) \leq \Lambda_0(\Lambda_\vartheta^{-1}(h_\vartheta(x) - \delta)) \mid X_i = x) < 1$$

since  $\Delta = h_0(x) - \Lambda_0(\Lambda_\vartheta^{-1}(\Lambda_\vartheta(\Lambda_0^{-1}(h_0(x))) - \delta)) > 0$  since each  $\Lambda_\vartheta$  is strictly increasing. The conditional distribution of  $\Lambda_\vartheta(Y_i)$  for some general  $\vartheta \in \Theta$  reads as

$$\mathbb{P}(\Lambda_\vartheta(Y_i) \leq y \mid X_i = x) = \mathbb{P}(\Lambda_0(Y_i) \leq \Lambda_0(\Lambda_\vartheta^{-1}(y)) \mid X_i = x) = F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y)) - h_0(x)).$$

**Remark 2.2** *It is important to give conditions under which the unknown components  $\Lambda_0 = \Lambda_{\vartheta_0}$ ,  $h_0 = h_{\vartheta_0}$  and  $F_0$  in model (2.1) are identifiable. To this end we impose the following assumptions.*

- *Assume that  $Y$  has a continuous distribution and w.l.o.g. assume that 0 is in the data range (otherwise shift the data).*
- *Assume that  $X$  is continuously distributed with support  $[0, 1]$ .*
- *Assume  $\Lambda_\vartheta(0) = 0$  for all  $\vartheta \in \Theta$ , and  $\Lambda_\vartheta$  is strictly increasing and continuous for each  $\vartheta \in \Theta$ .*
- *Assume that if for some  $\vartheta_0, \vartheta_1 \in \Theta$  one has*

$$(\Lambda_{\vartheta_1} \circ \Lambda_{\vartheta_0}^{-1})(a - b) = (\Lambda_{\vartheta_1} \circ \Lambda_{\vartheta_0}^{-1})(a) - (\Lambda_{\vartheta_1} \circ \Lambda_{\vartheta_0}^{-1})(b)$$

*for all  $a, b \in J$ , where  $J$  is an interval of positive length, then it follows that  $\vartheta_0 = \vartheta_1$ .*

- Assume that  $F_0$  (the cdf of  $\varepsilon = \Lambda_{\vartheta_0}(Y) - h_{\vartheta_0}(X)$ ) is strictly increasing.
- Assume that  $h_{\vartheta_0}$  is not constant and is continuous.

Now assume that the model

$$\Lambda_{\vartheta}(Y) = h_{\vartheta}(X) + \varepsilon(\vartheta) \text{ with } X \text{ independent from } \varepsilon(\vartheta)$$

holds for  $\vartheta = \vartheta_0$  (with our notations  $\Lambda_{\vartheta_0} = \Lambda_0$ ,  $h_{\vartheta_0} = h_0$ ,  $\varepsilon(\vartheta_0) = \varepsilon$ ) and for  $\vartheta = \vartheta_1$ .

Note that from the assumption it follows that  $\mathbb{P}(\varepsilon(\vartheta) \leq 0) = 1$ ,  $\mathbb{P}(\varepsilon(\vartheta) \leq -\Delta) < 1$  for each  $\Delta > 0$ , such that  $h_{\vartheta}$  is the upper boundary curve in the model (for  $\vartheta \in \{\vartheta_0, \vartheta_1\}$ ).

We show in section B of the appendix that it follows that  $\vartheta_0 = \vartheta_1$ . Thus the transformation is identifiable. Further  $h_{\vartheta_0}(x)$  is then the right endpoint of the conditional distribution of  $\Lambda_{\vartheta_0}(Y)$ , given  $X = x$ , and  $F_0$  is identified as cdf of  $\Lambda_{\vartheta_0}(Y) - h_{\vartheta_0}(X)$ .

If the function class  $\mathcal{L}$  contains the identity, then the assumptions rule out that it contains transformations which are linear on some interval with positive length. On the other hand it is clear that linear transformations can never reduce the dependence between the covariate and the error distribution.

**Example 2.3** In this example we give two classes of transformations that fulfill the identifiability assumptions.

Yeo and Johnson (2000) generalized the Box-Cox transformations by suggesting

$$\Lambda_{\vartheta}(y) = \begin{cases} \frac{(y+1)^{\vartheta}-1}{\vartheta}, & \text{if } y \geq 0, \vartheta \neq 0 \\ \log(y+1), & \text{if } y \geq 0, \vartheta = 0 \\ -\frac{(-y+1)^{2-\vartheta}-1}{2-\vartheta}, & \text{if } y < 0, \vartheta \neq 2 \\ -\log(-y+1), & \text{if } y < 0, \vartheta = 2, \end{cases}$$

which are typically considered for  $\vartheta \in \Theta = [0, 2]$  because then they are bijective maps  $\Lambda_{\vartheta} : \mathbb{R} \rightarrow \mathbb{R}$ . Note that  $\Lambda_{\vartheta}(0) = 0$  for all  $\vartheta \in \Theta$ .

The class of sinh-arcsinh transformations, see Jones and Pewsey (2009), do shift the location, but they can be modified to fulfill  $\Lambda_{\vartheta}(0) = 0$  for all  $\vartheta \in \Theta$ , e.g. consider

$$\Lambda_{(\vartheta_1, \vartheta_2)}(y) = \sinh(\vartheta_1 \sinh^{-1}(y) - \vartheta_2) - \sinh(-\vartheta_2).$$

Here  $\vartheta_1 > 0$  is the tailweight parameter and  $\vartheta_2 \in \mathbb{R}$  the skewness parameter. These transformations define also bijective maps  $\Lambda_{(\vartheta_1, \vartheta_2)} : \mathbb{R} \rightarrow \mathbb{R}$ .

## 2.2 The fixed design case

In the fixed design case we consider a triangular array of independent observations  $Y_{i,n}$ ,  $i = 1, \dots, n$ , and deterministic design points  $0 < x_{1,n} < \dots < x_{n,n} < 1$ . Once again we assume existence of a transformation  $\Lambda_0 = \Lambda_{\vartheta_0}$  in the class  $\mathcal{L}$  such that setting  $h_0 = h_{\vartheta_0}$  in the corresponding regression model

$$\Lambda_0(Y_{i,n}) = h_0(x_{i,n}) + \varepsilon_{i,n} \quad (2.2)$$

the cdf of the errors does not depend on the design points, i.e.  $\varepsilon_{i,n} \sim F_0 \forall i, n$ . Note that, as in the random design case, we assume  $F_0(0) = 1$  and  $F_0(-\Delta) < 1$  for all  $\Delta > 0$  leading again to  $h_{\vartheta} = \Lambda_{\vartheta} \circ \Lambda_0^{-1} \circ h_0$ .

**Remark 2.4** *Identifiability can be shown under the same conditions as in Remark 2.2 as long as  $\bar{\Delta}_n := \max_{1 \leq i \leq n+1} (x_{i,n} - x_{i-1,n}) \rightarrow 0$ ; see section D of the supplement.*

**Example 2.5** *Figures 1 and 2 show realizations of the original data and the transformed data (2.2) using a Yeo and Johnson transformation; see Example 2.3. For each figure, in the upper left panel the original data  $(x_{i,n}, Y_{i,n})$ ,  $i = 1, \dots, n = 100$ , are depicted with their boundary curve, while the upper right panel shows the corresponding non-iid errors. The lower left panel shows the transformed data with the curve  $h_0$ , while the lower right panel shows the iid errors  $\varepsilon_{i,n}$ ,  $i = 1, \dots, n$ .*

## 3 Estimating the transformation

### 3.1 The random design case

If  $\vartheta_0$  were known we could estimate the regression function (upper boundary curve)  $h_0$  by a local constant approximation, i.e.

$$\tilde{h}_0(x) = \max\{\Lambda_0(Y_i) | i = 1, \dots, n \text{ with } |X_i - x| \leq b_n\}, \quad (3.1)$$

where  $b_n \searrow 0$  is a sequence of bandwidths. For this estimator we will show uniform consistency under the following assumptions.

- (A1) Model (2.1) holds with iid  $\varepsilon_1, \dots, \varepsilon_n \sim F_0$  and  $F_0(0) = 1$ ,  $F_0(-\Delta) < 1$  for all  $\Delta > 0$ , and  $\varepsilon_1, \dots, \varepsilon_n$  are independent of  $X_1, \dots, X_n$ .
- (A2) The covariates  $X_1, \dots, X_n$  are iid with cdf  $F_X$  and density  $f_X$  that is continuous and bounded away from zero on its support  $[0, 1]$ .
- (A3) The regression function  $h_0$  is continuous on  $[0, 1]$ .

(A4) Let  $(b_n)_{n \in \mathbb{N}}$  be a sequence of positive bandwidths that satisfies  $\lim_{n \rightarrow \infty} b_n = 0$  and  $\lim_{n \rightarrow \infty} (\log n)/(nb_n) = 0$ .

Note that we do not require any assumption on the error distribution. In particular, in the setup of regularly varying distributed errors, all the results hold for regular as well as irregular distributions. In what follows, let  $\|\cdot\|_\infty$  denote the supremum norm and  $I\{\cdot\}$  the indicator function.

**Lemma 3.1** *Under model (2.1) with assumptions (A1)–(A4) we have  $\|\tilde{h}_0 - h_0\|_\infty = o_P(1)$ .*

The proof of the lemma is given in section A.1 of the appendix. The result applies for a model without transformation. Thus, as a by-product, we show uniform consistency of a boundary curve estimator in models with random covariates (and non-equidistant fixed design, see Lemma 3.3), while in contrast Drees et al. (2018) assumed equidistant design and obtained rates of convergence under stronger assumptions on the error distribution  $F_0$  and on the boundary curve  $h_0$ .

For general  $\vartheta \in \Theta$  we define a simple boundary curve estimator accordingly as

$$\tilde{h}_\vartheta(x) = \max\{\Lambda_\vartheta(Y_i) \mid i = 1, \dots, n \text{ with } |X_i - x| \leq b_n\}$$

and it holds that  $\tilde{h}_\vartheta = \Lambda_\vartheta \circ \Lambda_0^{-1} \circ \tilde{h}_0$ . Thus  $\tilde{h}_\vartheta$  consistently estimates  $h_\vartheta$ . The local constant estimator can be improved by introducing slight smoothing. To this end, let  $K$  be a density with compact support and  $a_n$  some sequence of bandwidths that decreases to zero such that  $na_n \rightarrow \infty$ . Define

$$\hat{h}_\vartheta(x) = \frac{\sum_{i=1}^n \tilde{h}_\vartheta(X_i) K\left(\frac{x-X_i}{a_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{a_n}\right)}, \quad (3.2)$$

then  $\hat{h}_\vartheta$  is also uniformly consistent for  $h_\vartheta$ ; see Lemma 4.2.

**Example 3.2** *For data as in Example 2.5, Figures 5 and 6 in the online supplementary material demonstrate the smoothing of the estimator. We use the Epanechnikov-kernel  $K(x) = 0.75(1 - x^2)I_{[-1,1]}(x)$  and bandwidths  $b_n = 0.5n^{-1/3}$ ,  $a_n = 0.5b_n$  with  $n = 100$ .*

Based on this estimator we define the joint empirical distribution function of residuals and covariates as  $\hat{F}_{n,\vartheta}(y, s) = \frac{1}{n} \sum_{i=1}^n I\{\Lambda_\vartheta(Y_i) - \hat{h}_\vartheta(X_i) \leq y\} I\{X_i \leq s\}$ . For  $\vartheta = \vartheta_0$ , the covariate  $X_i$  and the error  $\Lambda_\vartheta(Y_i) - h_\vartheta(X_i)$  are stochastically independent and thus, the joint empirical distribution function minus the product of the marginals, namely  $\hat{F}_{n,\vartheta}(y, s) - \hat{F}_{n,\vartheta}(y, 1)\hat{F}_{X,n}(s)$ , estimates zero for  $\vartheta = \vartheta_0$ . Here  $\hat{F}_{X,n}(\cdot) = \hat{F}_{n,\vartheta}(\infty, \cdot)$  denotes the empirical

distribution function of  $X_1, \dots, X_n$ . We will use this idea to estimate the transformation parameter  $\vartheta_0$ . To this end, for any function  $h : [0, 1] \rightarrow \mathbb{R}$  define

$$G_n(\vartheta, h)(y, s) = \frac{1}{n} \sum_{i=1}^n I\{\Lambda_\vartheta(Y_i) - h(X_i) \leq y\} (I\{X_i \leq s\} - \hat{F}_{X,n}(s)) \quad (3.3)$$

and note that  $G_n(\vartheta, \hat{h}_\vartheta)(y, s) = \hat{F}_{n,\vartheta}(y, s) - \hat{F}_{n,\vartheta}(y, 1)\hat{F}_{X,n}(s)$ . Our criterion function will be

$$M_n(\vartheta) = \|G_n(\vartheta, \hat{h}_\vartheta)\|$$

for some semi-norm  $\|\cdot\|$  as described in the following assumption.

**(N1)**  $\|\cdot\|$  is a semi-norm such that  $\|\Gamma\| \leq c \sup_{\substack{y \in C \\ s \in [0,1]}} |\Gamma(y, s)|$  for some constant  $c > 0$  and some compact set  $C = [c_1, c_2] \subset \mathbb{R}$  with  $c_1, c_2 > 0$  and  $0 \in C$ , for all measurable functions  $\Gamma : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R}$ .

For instance one can consider one of the following semi-norms,

- (i)  $\|\Gamma(y, s)\| = \sup_{\substack{s \in [0,1] \\ y \in C}} |\Gamma(y, s)|$
- (ii)  $\|\Gamma(y, s)\| = \left( \int \Gamma(y, s)^2 w(y, s) d(y, s) \right)^{1/2}$  for some integrable weight function  $w : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R}_0^+$  with support included in  $C \times [0, 1]$
- (iii)  $\|\Gamma(y, s)\| = \sup_{s \in [0,1]} \left( \int \Gamma(y, s)^2 w(y) dy \right)^{1/2}$  for some integrable weight function  $w : \mathbb{R} \rightarrow \mathbb{R}_0^+$  with support included in  $C$
- (iv)  $\|\Gamma(y, s)\| = \sup_{y \in C} \left( \int \Gamma(y, s)^2 w(s) ds \right)^{1/2}$  for some integrable weight function  $w : [0, 1] \rightarrow \mathbb{R}_0^+$ .

The first two semi-norms correspond to Kolmogorov-Smirnov and Cramér-von Mises distances, respectively, while the last two are mixtures of both.

Now we define the estimator  $\hat{\vartheta}$  of  $\vartheta_0$  as the minimizer of  $M_n(\vartheta)$  over  $\Theta$ , i.e.

$$\hat{\vartheta} = \arg \min_{\vartheta \in \Theta} M_n(\vartheta). \quad (3.4)$$

For the following theory also the weaker condition  $M_n(\hat{\vartheta}) \leq \inf_{\vartheta \in \Theta} M_n(\vartheta) + o_P(1)$  is sufficient. Note that

$$\mathbb{P}(\Lambda_\vartheta(Y_i) - h(X_i) \leq y \mid X_i = x) = \mathbb{P}(\Lambda_0(Y_i) \leq \Lambda_0(\Lambda_\vartheta^{-1}(y + h(x))) \mid X_i = x)$$

$$= F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y + h(x))) - h_0(x)),$$

which reduces to  $F_0(y)$  for  $\vartheta = \vartheta_0$  and  $h = h_0$ . Now considering expectations we define

$$\begin{aligned} G(\vartheta, h)(y, s) &= \int F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y + h(x))) - h_0(x)) I\{x \leq s\} f_X(x) dx \\ &\quad - \int F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y + h(x))) - h_0(x)) f_X(x) dx F_X(s) \end{aligned} \quad (3.5)$$

as deterministic counterpart of  $G_n(\vartheta, h)$ . Further set

$$M(\vartheta) = \|G(\vartheta, h_\vartheta)\|$$

and note that  $M(\vartheta_0) = \|G(\vartheta_0, h_0)\| = 0$ . In section 4 we formulate assumptions under which  $\hat{\vartheta}$  consistently estimates  $\vartheta_0$ .

### 3.2 The fixed design case

In the fixed design model (2.2) we define the estimator for the boundary curve  $h_0$  as

$$\tilde{h}_0(x) = \max\{\Lambda_0(Y_{i,n}) | i = 1, \dots, n \text{ with } |x_{i,n} - x| \leq b_n\}$$

and obtain uniform consistency under the following modified assumptions. We set  $x_{0,n} = 0$  and  $x_{n+1,n} = 1$ .

**(A1')** Model (2.2) holds with independent  $\varepsilon_{1,n}, \dots, \varepsilon_{n,n}$  with cdf  $F_0$  ( $\forall n$ ) such that  $F_0(0) = 1$ ,  $F_0(-\Delta) < 1$  for all  $\Delta > 0$ .

**(A2')** The design points  $0 < x_{1,n} < \dots < x_{n,n} < 1$  are deterministic.

**(A4')** Let  $(b_n)_{n \geq 0}$  be a sequence of positive bandwidths that satisfies  $\lim_{n \rightarrow \infty} b_n = 0$  and  $\lim_{n \rightarrow \infty} \bar{\Delta}_n \log(n)/b_n = 0$  for  $\bar{\Delta}_n := \max_{1 \leq i \leq n+1} (x_{i,n} - x_{i-1,n})$ .

**Lemma 3.3** *Under model (2.2) with assumptions (A1'), (A2'), (A3) and (A4') we have  $\|\tilde{h}_0 - h_0\|_\infty = o_P(1)$ .*

The proof is given in section C.1 of the online supplementary material. For general  $\vartheta \in \Theta$  we define a consistent boundary curve estimator as

$$\tilde{h}_\vartheta(x) = \max\{\Lambda_\vartheta(Y_{i,n}) | i = 1, \dots, n \text{ with } |x_{i,n} - x| \leq b_n\} = \Lambda_\vartheta(\Lambda_0^{-1}(\tilde{h}_0(x))).$$

In analogy to (3.3) we define, for any function  $h : [0, 1] \rightarrow \mathbb{R}$ ,

$$G_n(\vartheta, h)(y, s) = \frac{1}{n} \sum_{i=1}^n I\{\Lambda_\vartheta(Y_{i,n}) - h(x_{i,n}) \leq y\} (I\{x_{i,n} \leq s\} - \hat{F}_{X,n}(s)), \quad (3.6)$$

where

$$\hat{F}_{X,n}(s) = \frac{1}{n} \sum_{i=1}^n I\{x_{i,n} \leq s\}.$$

The criterion function is again  $M_n(\vartheta) = \|G_n(\vartheta, \hat{h}_\vartheta)\|$  where the smooth estimator  $\hat{h}_\vartheta$  is defined accordingly as in (3.2) and with this the transformation parameter estimator is similar to (3.4). In order to consider the same deterministic  $G$  as in (3.5) an additional assumption is needed.

**(A2'')** The design points  $0 < x_{1,n} < \dots < x_{n,n} < 1$  are deterministic. There exists a cdf  $F_X$  with continuous density function  $f_X : [0, 1] \rightarrow \mathbb{R}$  which is bounded away from zero such that

$$\max_{i=1, \dots, n+1} \left| \int_{x_{i-1,n}}^{x_{i,n}} f_X(x) dx - \frac{1}{n} \right| = o\left(\frac{1}{n}\right).$$

Assumption **(A2'')** is common in the literature on fixed design regression models. It allows the application of the mean value theorem for integrals to obtain, for some  $\xi_{i,n} \in [x_{i-1,n}, x_{i,n}]$ ,

$$f_X(\xi_{i,n})(x_{i,n} - x_{i-1,n}) = \int_{x_{i-1,n}}^{x_{i,n}} f_X(x) dx = \frac{1}{n} + o\left(\frac{1}{n}\right)$$

uniformly in  $i = 1, \dots, n$ . Thus it follows from **(A2'')** that  $\bar{\Delta}_n$  in assumption **(A4')** has the exact rate  $n^{-1}$  and therefore assumption **(A4')** reduces to **(A4)**. Further the following Riemann sum approximations for bounded integrable functions  $\varphi$  can be applied to get

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \varphi(x_{i,n}) &= \sum_{i=1}^n \varphi(x_{i,n}) (f_X(\xi_{i,n})(x_{i,n} - x_{i-1,n}) + o(\frac{1}{n})) \\ &= \int \varphi(x) f_X(x) dx + o(1). \end{aligned} \tag{3.7}$$

In the next section we state conditions under which  $\hat{\vartheta} = \arg \min_{\vartheta \in \Theta} M_n(\vartheta)$  consistently estimates  $\vartheta_0$ .

## 4 Main result

To prove consistency of the estimator for the transformation parameter we need the following additional assumptions. Please note that assumption **(B1)** implies identifiability of the transformation  $\Lambda_0$  in the class  $\mathcal{L}$ .

**(B1)** For every  $\delta > 0$  there exists some  $\epsilon > 0$  such that  $\inf_{\|\vartheta - \vartheta_0\| > \delta} M(\vartheta) \geq \epsilon$ .

**(B2)**  $\mathcal{L} = \{\Lambda_\vartheta \mid \vartheta \in \Theta\}$  is a class of strictly increasing continuous functions  $\mathbb{R} \rightarrow \mathbb{R}$ .

- (B3)** Let  $S = \{\Lambda_0^{-1}(h_0(x)) \mid x \in [0, 1]\}$ . Then the class  $\mathcal{L}_S = \{\Lambda_\vartheta|_S \mid \vartheta \in \Theta\}$  is pointwise bounded and uniformly equicontinuous, i.e.  $\sup_{\vartheta \in \Theta} |\Lambda_\vartheta(y)| < \infty$  for all  $y \in S$ , and for every  $\epsilon > 0$  there exists some  $\delta > 0$  such that  $\sup_{\vartheta \in \Theta} |\Lambda_\vartheta(y) - \Lambda_\vartheta(z)| < \epsilon$  for all  $y, z \in S$  with  $|y - z| \leq \delta$ .
- (B4)** The class  $\mathcal{L}_{\tilde{S}}^1 = \{\Lambda_0 \circ \Lambda_\vartheta^{-1}|_{\tilde{S}} \mid \vartheta \in \Theta\}$  is pointwise bounded and uniformly equicontinuous for  $\tilde{S} = \{z + h_\vartheta(x) \mid z \in C_\tau, \vartheta \in \Theta, x \in [0, 1]\}$  with  $C_\tau = [c_1 - \tau, c_2 + \tau]$  (for  $C = [c_1, c_2]$  from **(N1)**) for some  $\tau > 0$ , i.e.  $\sup_{\vartheta \in \Theta} |\Lambda_0(\Lambda_\vartheta^{-1}(z))| < \infty$  for all  $z \in \tilde{S}$ , and for every  $\delta > 0$  there exists some  $\gamma > 0$  such that  $\sup_{\vartheta \in \Theta} |\Lambda_0(\Lambda_\vartheta^{-1}(x)) - \Lambda_0(\Lambda_\vartheta^{-1}(z))| \leq \delta$  for all  $x, z \in \tilde{S}$  with  $|x - z| \leq \gamma$ .
- (B5)** For some  $\tau > 0$ ,  $F_0$  is uniformly continuous on the set  $\tilde{C} = \{\Lambda_0(\Lambda_\vartheta^{-1}(y + a + h_\vartheta(x))) - h_0(x) \mid y \in C, \vartheta \in \Theta, x \in [0, 1], |a| \leq \tau\}$  (with  $C$  from **(N1)**), i.e. for every  $\epsilon > 0$  there is some  $\delta > 0$  such that  $|F_0(y) - F_0(z)| < \epsilon$  if  $|y - z| \leq \delta$ ,  $y, z \in \tilde{C}$ .
- (B6)**  $K$  is a density with support  $[-1, 1]$  and  $b_n \searrow 0$ ,  $nb_n \rightarrow \infty$ .

Let us now make few comments regarding these assumptions.

- **(B1)** is a common assumption in M-estimation and needed for uniqueness of the true parameter.
- **(B2)** implies the existence of continuous inverse functions  $\Lambda_\vartheta^{-1}$ . Further note that uniform equicontinuity and pointwise boundedness imply totally boundedness by the Arzelà-Ascoli theorem. Thus for each  $\epsilon$  there is a finite covering of the classes  $\mathcal{L}_S$  from **(B3)** and  $\mathcal{L}_{\tilde{S}}^1$  from **(B4)** with balls of radius  $\epsilon$  with respect to the sup norm. Thus also the sup norm bracketing numbers of those classes are finite, i.e.

$$N_{[]}(\epsilon, \mathcal{L}_S, \|\cdot\|_\infty) < \infty, \quad N_{[]}(\epsilon, \mathcal{L}_{\tilde{S}}^1, \|\cdot\|_\infty) < \infty \text{ for all } \epsilon > 0 \quad (4.1)$$

(see, e.g., Lemma 9.21 in Kosorok, 2008).

- **(B3)**–**(B5)** can be seen as minimal assumptions on the class  $\mathcal{L} = \{\Lambda_\vartheta \mid \vartheta \in \Theta\}$  and  $F_0$ . As typically the sets  $S$ ,  $\tilde{S}$  and  $\tilde{C}$  are unknown, the assumptions can be replaced by stronger assumptions that hold on all compact sets. Besides, working on compact set transformation parameter, assumptions **(B3)**–**(B4)** hold for most of transformations used in practice such as the Box and Cox transformations (see Box and Cox, 1964) (suitably modified taking into account the data range), the exponential transformations (see Manly, 1976), the sinh-arcsinh transformations (see Jones and Pewsey, 2009). For instance, with regard to Yeo-Johnson transformations, when  $\vartheta \in \Theta = [0, 2]$ ,  $\Lambda_\vartheta : \mathbb{R} \rightarrow \mathbb{R}$  defines a bijective map (see Remark 2.3) and both  $\Lambda_\vartheta$  and  $\Lambda_\vartheta^{-1}$  have uniform bounded

derivatives on compact sets so that one may show that they fulfill assumptions **(B3)**–**(B4)** using the mean value theorem to  $\Lambda_\vartheta$  and  $\Lambda_\vartheta^{-1}$ . Further under stronger assumptions on the smoothness of  $F_0$ ,  $\Lambda_\vartheta$  and  $\Lambda_0 \circ \Lambda_\vartheta^{-1}$ , the theoretical results can be generalized to semi-norms that are not restricted to a compact  $C \times [0, 1]$  as in assumption **(N1)**.

- **(B6)** is standard in kernel smoothing and is needed for the smoothed estimator  $\hat{h}_\vartheta$  to be consistent. While we noticed in the simulations that slight smoothing improves the procedure, the following theorem still holds when  $\hat{h}_\vartheta$  is replaced by the non-smooth estimator  $\tilde{h}_\vartheta$ . Assumption **(B5)** holds, e.g. for Hölder-continuous distribution functions  $F_0$ .

The following theorem states consistency of the transformation parameter estimator.

**Theorem 4.1 (i).** *(The random design case.) Assume model (2.1) under assumptions **(A1)**–**(A4)**, **(N1)**, **(B1)**–**(B6)**. Then  $\hat{\vartheta}$  is a consistent estimator, i.e.  $\hat{\vartheta} - \vartheta_0 = o_P(1)$ .*

**(ii).** *(The fixed design case.) Assume model (2.2) under assumptions **(A1')**, **(A2')**, **(A3)**, **(A4)**, **(N1)**, **(B1)**–**(B6)**. Then  $\hat{\vartheta}$  is a consistent estimator, i.e.  $\hat{\vartheta} - \vartheta_0 = o_P(1)$ .*

The proof for the random design case is given in section A.3 of the appendix and the proof for the fixed design case in section C.2 of the supplement. One basic ingredient is the following result, which is proven in section A.2 of the appendix for the random design case. The proof for the fixed design case is analogous.

**Lemma 4.2 (i).** *(The random design case.) Under model (2.1) with assumptions **(A1)**–**(A4)**, **(B2)**, **(B3)**, **(B6)**, we have  $\sup_{\vartheta \in \Theta} \|\hat{h}_\vartheta - h_\vartheta\|_\infty = o_P(1)$ .*

**(ii).** *(The fixed design case.) Under model (2.2) with assumptions **(A1')**, **(A2')**, **(A3)**, **(A4')**, **(B2)**, **(B3)**, **(B6)**, we have  $\sup_{\vartheta \in \Theta} \|\hat{h}_\vartheta - h_\vartheta\|_\infty = o_P(1)$ .*

The consistency result in Theorem 4.1 should be seen as a first step in the analysis of transformation boundary regression models. An interesting and challenging topic for future research is to derive an asymptotic distribution of  $\hat{\vartheta} - \vartheta_0$  (properly scaled) and to investigate the asymptotic influence of the estimation on subsequent procedures based on the transformed data. This is beyond the scope of the paper as yet there are no results on the uniform asymptotic distribution of  $\tilde{h}_0 - h_0$  in the literature.

We finally highlight that under the further condition **(A3')** defined below regarding the regularity of the boundary curve, we obtain as a corollary of Theorem 4.1 the consistency of the estimator  $\hat{h}_\vartheta$  of the boundary curve.

(A3')  $\vartheta_0$  is an inner point of a convex parameter space  $\Theta$  and  $h_\vartheta$  is continuously differentiable with respect to  $\vartheta$ . Besides, we assume that there exists some  $\delta > 0$  such that

$$\sup_{x \in [0,1]} \sup_{\|\vartheta - \vartheta_0\| < \delta} \left\| \frac{\partial h_\vartheta(x)}{\partial \vartheta} \right\| < \infty.$$

**Corollary 4.3 (i).** *(The random design case.) Assume model (2.1) holds under assumptions (A1), (A2), (A3'), (A4), (N1) and (B1)–(B6). Then  $\hat{h}_{\hat{\vartheta}}$  is a consistent estimator of  $h_{\vartheta_0}$ , i.e.  $\|\hat{h}_{\hat{\vartheta}} - h_{\vartheta_0}\|_\infty = o_P(1)$ .*

**(ii).** *(The fixed design case.) Assume model (2.2) holds under assumptions (A1'), (A2''), (A3'), (A4), (N1) and (B1)–(B6). Then  $\hat{h}_{\hat{\vartheta}}$  is a consistent estimator of  $h_{\vartheta_0}$ , i.e.  $\|\hat{h}_{\hat{\vartheta}} - h_{\vartheta_0}\|_\infty = o_P(1)$ .*

**Proof.** We only prove (i) since the proof of (ii) is identical. Observe first that

$$\|\hat{h}_{\hat{\vartheta}} - h_{\vartheta_0}\|_\infty \leq \|\hat{h}_{\hat{\vartheta}} - h_{\hat{\vartheta}}\|_\infty + \|h_{\hat{\vartheta}} - h_{\vartheta_0}\|_\infty.$$

The first term in the right hand side of the above inequality goes to 0 in probability from Lemma 4.2 since the consistency holds uniformly over  $\vartheta \in \Theta$ . Regarding the second term, applying the mean value theorem, there exists some  $\vartheta^*(x)$  on the line between  $\hat{\vartheta}$  and  $\vartheta_0$  such that

$$\|h_{\hat{\vartheta}} - h_{\vartheta_0}\|_\infty = \sup_{x \in [0,1]} \left| \frac{\partial h_\vartheta(x)^T}{\partial \vartheta} \Big|_{\vartheta = \vartheta^*(x)} (\hat{\vartheta} - \vartheta_0) \right|.$$

From Theorem 4.1,  $\hat{\vartheta} - \vartheta_0 = o_P(1)$  which concludes the proof under the assumption (A3').

□

## 5 Simulations

To study the small sample behavior, we generate data as  $Y = \Lambda_{\vartheta_0}^{-1}(h_0(x) + \varepsilon)$  using the Yeo-Johnson transformation for different values of  $\vartheta_0$ . We focus on the equidistant design framework and examine the two regression functions  $h_0(x) = 10(x - \frac{1}{2})^2$  and  $h_0(x) = \frac{1}{2} \sin(2\pi x) + 4x$  for two different error distributions, namely the Weibull distribution with scale parameter 1 and shape parameter 3 and the exponential distribution with mean 1/3. We consider samples of size  $n = 50$  and  $n = 100$ . It means that we investigate the following four models

$$h_0(x) = 10(x - \frac{1}{2})^2 \quad \text{with} \quad \varepsilon \sim \text{Weibull}(1, 3) \quad (5.1)$$

$$h_0(x) = 10(x - \frac{1}{2})^2 \quad \text{with} \quad \varepsilon \sim \text{Exp}(3) \quad (5.2)$$

$$h_0(x) = \frac{1}{2} \sin(2\pi x) + 4x \quad \text{with} \quad \varepsilon \sim \text{Weibull}(1, 3) \quad (5.3)$$

$$h_0(x) = \frac{1}{2} \sin(2\pi x) + 4x \quad \text{with} \quad \varepsilon \sim \text{Exp}(3). \quad (5.4)$$

Figures 1 and 2 show realizations of models (5.1) and (5.3). The bandwidth  $b_n = n^{-1/3}$  is chosen accordingly to Drees et al. (2018) and simulations are based on 1000 iterations. We use the Epanechnikov kernel to smooth the boundary curve estimator and compare the results for two smoothing parameters  $a_n = b_n/2$  and  $a_n = b_n/20$ . The transformation parameter estimator is as in (3.4) on the interval  $[-0.5, 2.5]$ , where the semi-norm in the criterion function  $M_n(\vartheta)$  is chosen as in (i), (ii), (iii), and (iv) in the examples of Condition **(N1)**. In the following we denote the according estimators as TKS, TCM, TKSCM and TCMKS. Here, TKS and TCM refer to Kolmogorov-Smirnov and Cramér-von Mises distances respectively, while TKSCM and TCMKS are mixtures of both. For simplicity, the weight functions are chosen identically equal to 1 in all the settings, i.e.,  $w(y, s) = 1$  for all  $(y, s) \in \mathbb{R} \times [0, 1]$ ,  $w(y) = 1$  for all  $y \in \mathbb{R}$  and  $w(s) = 1$  for all  $s \in [0, 1]$  in (ii), (iii) and (iv), respectively (although for the theory we assumed a compact support).

We sum up the simulation results in the following 8 tables. Tables 1, 3 and Tables 2, 4 deal with Models (5.1) and (5.4) for  $a_n = b_n/2$  and  $a_n = b_n/20$ , respectively, whereas Tables 6, 8 and Tables 7, 9 in the supplement show the results for Models (5.2) and (5.3). In Figure 3, we have represented the density function of each estimator for the Model (5.1) when  $\vartheta_0 = 0.5$  with  $n = 100$  and  $a_n = b_n/20$ , which corresponds to the settings of Table 2. To assess the performance of our estimates, we provide for each estimator the mean, the median and the Mean Integrated Squared Error (MISE) in brackets for five values of the true parameter  $\vartheta_0 = 0, 0.5, 1, 1.5, 2$ . The best-performing one regarding the mean (respectively the MISE) is highlighted in bold (respectively is underlined).

Looking at the MISE, it turns out that the estimator using the Cramér-von Mises distance (TCM) out-performs in many cases even when it does not out-perform the mean; see Tables 1 and 2 when  $n = 100$  for instance. Besides, as it is intended, results are better in most of the cases when the sample size  $n$  increases. However, this does not hold for every case. For instance, one may see in Table 2 that for the second estimator TCM, most of the results are better for  $n = 50$  than for  $n = 100$ . This might relate to a sensitivity with respect to the choices of bandwidth and smoothing parameter. A lot of criteria may be used to judge the performance of the estimators. We deal here with the mean, the median and the MISE but we emphasize that using different criteria (e.g. median absolute deviation, mode or even graphical analysis) could give different results concerning the comparison of the methods. For instance, results in Table 3 for  $n = 100$ ,  $a_n = b_n/2$  and  $\vartheta_0 = 1$  are quite not accurated regarding the mean (e.g. 0.845 for the TCM). Nevertheless, looking at Figure 4, it appears that the plots of the densities look satisfactory.

It is clear that the TCM and the TKSCM out-perform in the Model (5.3) and in the Model (5.4), respectively. Nonetheless, in a general setting, we are not able to state which estimator

performs better since it depends first on the criterium selected to judge the performance but more importantly on the choice of the bandwidths and the smoothing parameter.

Finally, we recall that the aim of this work is to reduce the dependence between the covariates and the errors. As one can see in Table 5, although the estimation of  $\vartheta_0$  is less good than expected in the model (5.4), the correlations between the covariates and the errors (after transformation) are very small; see also Table 10 in the supplement for the correlations in the model (5.3). We obtain similar results for the random design case.

## A Proofs of asymptotic results in the random covariate case

For the proofs of the asymptotic results let us fix some notation:  $\lfloor \cdot \rfloor$  and  $\lceil \cdot \rceil$  are the floor and ceiling functions respectively;  $\bar{F} = 1 - F$  denotes the survival function associated to a cdf  $F$ ;  $X_1 \stackrel{d}{=} X_2$  means that two random variables  $X_1, X_2$  share the same distribution;  $a_n \underset{n \rightarrow \infty}{\sim} b_n$  holds if there exists a constant  $c > 0$  such that  $\lim_{n \rightarrow \infty} a_n/b_n = c$  for two sequences  $(a_n)_{n \geq 1}$  and  $(b_n)_{n \geq 1}$  of nonnegative numbers;  $A^c$  is the complement of a set  $A$ .

In the following we give the proofs of our results in the random design case whereas the proofs for the fixed design case can be found in the online supplementary material.

### A.1 Proof of Lemma 3.1

At first we need the following intermediary lemma.

**Lemma A.1** *Assume model (2.1) holds with assumptions (A1), (A2) and (A4). Then, we have*

$$\sup_{x \in [0,1]} \min_{\substack{i \in \{1, \dots, n\} \\ |X_i - x| \leq b_n}} |\varepsilon_i| = o_P(1).$$

**Proof.** For  $n \geq 1$  denote  $X_{(1)} < X_{(2)} < \dots < X_{(n)}$  the order statistics of the random design sample  $X_1, X_2, \dots, X_n$ . Let  $\pi$  be the random permutation of  $\{1, \dots, n\}$  such that  $X_{(i)} = X_{\pi(i)}, i = 1, \dots, n$ . Due to the independence between the errors and the covariates under (A1),  $\varepsilon_{\pi(1)}, \dots, \varepsilon_{\pi(n)}$  are iid with cdf  $F_0$ . Let  $Z_i = -\varepsilon_{\pi(i)}, i = 1, \dots, n$ , then  $Z_1, \dots, Z_n$  are iid with cdf  $U$  with  $U(x) = 1 - F_0(-x)$  and we need to show that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \sup_{x \in [0,1]} \min_{\substack{i \in \{1, \dots, n\} \\ |X_{(i)} - x| \leq b_n}} Z_i > \epsilon \right) = 0, \quad \epsilon > 0. \quad (\text{A.1})$$

Define for  $n \geq 1$  the event

$$\Omega_n = \left\{ \inf_{x \in [0,1]} \sum_{i=1}^n I\{|X_i - x| \leq b_n\} \geq Cnb_n \right\}$$

for a suitable constant  $C > 0$  specified later. Note that on  $\Omega_n$ , there are at least  $Cnb_n$  covariates in each of the intervals  $[x - b_n, x + b_n]$ . We will first show that  $\lim_{n \rightarrow \infty} \mathbb{P}(\Omega_n) = 1$ . To this end, for  $n \geq 1$  let  $f_{n,x}(z) = I\{|x - z| \leq b_n\}$  and note that

$$\begin{aligned} \inf_{x \in [0,1]} \frac{1}{n} \sum_{i=1}^n I\{|X_i - x| \leq b_n\} &\geq \inf_{x \in [0,1]} \mathbb{P}(|X_1 - x| \leq b_n) \\ &\quad - \sup_{x \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^n (f_{n,x}(X_i) - \mathbb{E}[f_{n,x}(X_i)]) \right|. \end{aligned} \quad (\text{A.2})$$

Applying the mean value theorem of integration, it follows that

$$2b_n \sup_{x \in [0,1]} f_X(x) \geq \mathbb{P}(|X_1 - x| \leq b_n) = \int_{\max(0, x-b_n)}^{\min(1, x+b_n)} f_X(x) dx \geq b_n \inf_{x \in [0,1]} f_X(x). \quad (\text{A.3})$$

Then, there exists a constant  $C_1 > 0$ , which actually corresponds to the lower bound of the density function  $f_X$  involved in Assumption **(A2)** such that

$$\mathbb{P}(|X_1 - x| \leq b_n) \geq C_1 b_n, \quad (\text{A.4})$$

uniformly over  $x \in [0, 1]$ .

Fix  $n \geq 1$  and denote  $P_n f_{n,x} := \frac{1}{n} \sum_{i=1}^n f_{n,x}(X_i)$  and  $P f_{n,x} := \mathbb{E}[f_{n,x}(X_1)]$  so that  $P_n$  and  $P$  refer to the empirical measure and the distribution of the random design sample  $X_1, \dots, X_n$ , respectively. By (A.3)  $P f_{n,x}^2 = \mathbb{E}[I\{|X - x| \leq b_n\}] \leq 2C_2 b_n$ , where  $C_2 := \sup_{x \in [0,1]} f_X(x)$ , which is finite under **(A2)**. Moreover, since  $|f_{n,x}(X)| \leq 1$  and the assumption on the covering number is fulfilled (see Example 38 and Problem 28 to be convinced in Pollard (1984)), Theorem 37 in Pollard (1984, p. 34) holds and we have

$$\sup_{x \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^n (f_{n,x}(X_i) - \mathbb{E}[f_{n,x}(X_i)]) \right| = o(b_n).$$

From this together with (A.2) and (A.4) it follows that  $\lim_{n \rightarrow \infty} \mathbb{P}(\Omega_n) = 1$ . It means that for any sub-interval  $I_n := [x - b_n, x + b_n]$ , there are at least  $Cnb_n$  random design points with probability converging to 1.

Then, for all  $y > 0$ , we have with  $d_n := \lceil Cnb_n \rceil$

$$\mathbb{P} \left( \sup_{x \in [0,1]} \min_{\substack{i \in \{1, \dots, n\} \\ |X_{(i)} - x| \leq b_n}} Z_i > y \right) \leq \mathbb{P} \left( \left\{ \max_{j \in \{1, \dots, n-d_n\}} \min_{i \in \{j, \dots, j+d_n\}} Z_i > y \right\} \cap \Omega_n \right) + \mathbb{P}(\Omega_n^c)$$

$$\begin{aligned}
&\leq \sum_{j=1}^{n-d_n} \mathbb{P} \left( \min_{i \in \{j, \dots, j+d_n\}} Z_i > y \right) + \mathbb{P}(\Omega_n^c) \\
&= (n-d_n) \mathbb{P} \left( \min_{i \in \{1, \dots, d_n+1\}} Z_i > y \right) + \mathbb{P}(\Omega_n^c) \\
&= (n-d_n) \bar{U}(y)^{d_n+1} + \mathbb{P}(\Omega_n^c).
\end{aligned}$$

Thus it remains to show that for all  $\epsilon > 0$

$$(n-d_n) \bar{U}(\epsilon)^{d_n+1} \xrightarrow{n \rightarrow \infty} 0$$

which is true since  $d_n \underset{n \rightarrow \infty}{\sim} nb_n$  and

$$\begin{aligned}
nb_n \log(\bar{U}(\epsilon)) + \log(n-d_n) &\leq nb_n \log(\bar{U}(\epsilon)) + \log(n) \\
&= \log(n) \left( \frac{nb_n}{\log(n)} \log(\bar{U}(\epsilon)) + 1 \right) \\
&\xrightarrow{n \rightarrow \infty} -\infty
\end{aligned}$$

since  $\bar{U}(\epsilon) < 1$  under **(A1)** and  $\frac{nb_n}{\log(n)} \xrightarrow{n \rightarrow \infty} \infty$  under **(A4)**. This concludes the proof.  $\square$

We are now ready to prove Lemma 3.1.

**Proof of Lemma 3.1.** On the one hand, we have

$$\begin{aligned}
\sup_{x \in [0,1]} \left( \tilde{h}_0(x) - h_0(x) \right) &= \sup_{x \in [0,1]} \left( \max_{\substack{i \in \{1, \dots, n\} \\ |X_i - x| \leq b_n}} \{h_0(X_i) + \varepsilon_i - h_0(x)\} \right) \\
&\leq \sup_{x \in [0,1]} \left( \max_{\substack{i \in \{1, \dots, n\} \\ |X_i - x| \leq b_n}} \{h_0(X_i) - h_0(x)\} \right) \\
&\leq \sup_{|t-x| \leq b_n} |h_0(t) - h_0(x)| \\
&= o(1),
\end{aligned} \tag{A.5}$$

since the errors  $(\varepsilon_i)_{1 \leq i \leq n}$  are nonpositive and  $h_0$  is continuous on the compact set  $[0, 1]$  and thereby uniformly continuous under **(A3)**. On the other hand,

$$\begin{aligned}
\sup_{x \in [0,1]} \left( h_0(x) - \tilde{h}_0(x) \right) &= \sup_{x \in [0,1]} \left( h_0(x) - \max_{\substack{i \in \{1, \dots, n\} \\ |X_i - x| \leq b_n}} \{h_0(X_i) + \varepsilon_{i,n}\} \right) \\
&= \sup_{x \in [0,1]} \left( \min_{\substack{i \in \{1, \dots, n\} \\ |X_i - x| \leq b_n}} \{h_0(x) - h_0(X_i) - \varepsilon_{i,n}\} \right) \\
&\leq \sup_{x \in [0,1]} \left( \min_{\substack{i \in \{1, \dots, n\} \\ |X_i - x| \leq b_n}} \{-\varepsilon_i\} \right) + \sup_{|t-x| \leq b_n} |h_0(t) - h_0(x)|
\end{aligned} \tag{A.6}$$

$$= o_P(1),$$

with Lemma A.1. Finally, combining equations (A.5) and (A.6), it follows that

$$\begin{aligned} \|h_0 - \tilde{h}_0\|_\infty &= \sup_{x \in [0,1]} \left| \tilde{h}_0(x) - h_0(x) \right| \\ &= \sup_{x \in [0,1]} \left( \max \left\{ \tilde{h}_0(x) - h_0(x), h_0(x) - \tilde{h}_0(x) \right\} \right) \\ &\leq \max \left\{ \sup_{x \in [0,1]} (\tilde{h}_0(x) - h_0(x)), \sup_{x \in [0,1]} (h_0(x) - \tilde{h}_0(x)) \right\} = o_P(1), \end{aligned}$$

which is the desired result.  $\square$

## A.2 Proof of Lemma 4.2

Let  $\epsilon > 0$ . Note that  $\Lambda_0^{-1} \circ h_0$  is uniformly continuous due to assumptions **(A3)** and **(B2)**. Thus with  $h_\vartheta = \Lambda_\vartheta \circ \Lambda_0^{-1} \circ h_0$  and assumption **(B3)** it follows that there exists some  $\delta > 0$  such that  $\sup_{\vartheta \in \Theta} |h_\vartheta(x) - h_\vartheta(y)| \leq \epsilon$  if  $|x - y| \leq \delta$ . Now let  $n$  be large enough such that  $b_n \leq \delta$ . Then due to the definition of  $\hat{h}_\vartheta$  and  $\text{supp}(K) = [-1, 1]$  one obtains

$$\|\hat{h}_\vartheta - h_\vartheta\|_\infty \leq \|\tilde{h}_\vartheta - h_\vartheta\|_\infty + \sup_{x \in [0,1]} \left| \frac{\sum_{i=1}^n (h_\vartheta(X_i) - h_\vartheta(x)) K\left(\frac{x-X_i}{b_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{b_n}\right)} \right| \leq \|\tilde{h}_\vartheta - h_\vartheta\|_\infty + \epsilon.$$

From Lemma 3.1 we have  $\|\tilde{h}_0 - h_0\|_\infty = o_P(1)$  and thus with assumption **(B3)** it follows that

$$\sup_{\vartheta \in \Theta} \|\tilde{h}_\vartheta - h_\vartheta\|_\infty = \sup_{\vartheta \in \Theta} \sup_{x \in [0,1]} |\Lambda_\vartheta(\Lambda_0^{-1}(\tilde{h}_0(x))) - \Lambda_\vartheta(\Lambda_0^{-1}(h_0(x)))| = o_P(1)$$

and therefore the assertion of the lemma.  $\square$

## A.3 Proof of Theorem 4.1

By the argmax theorem applied to the criterion function  $M_n(\vartheta)$  multiplied by  $(-1)$  and using assumption **(B1)** it suffices to show that

$$\sup_{\vartheta \in \Theta} |M_n(\vartheta) - M(\vartheta)| = o_P(1)$$

(see Kosorok, 2008, Theorem 2.12(i)). To obtain this, note that

$$\sup_{\vartheta \in \Theta} |M_n(\vartheta) - M(\vartheta)| \leq \sup_{\vartheta \in \Theta} \|G_n(\vartheta, \hat{h}_\vartheta) - \bar{G}_n(\vartheta, \hat{h}_\vartheta)\| + \sup_{\vartheta \in \Theta} \|\bar{G}_n(\vartheta, \hat{h}_\vartheta) - G(\vartheta, \hat{h}_\vartheta)\|$$

$$+ \sup_{\vartheta \in \Theta} \|G(\vartheta, \hat{h}_\vartheta) - G(\vartheta, h_\vartheta)\|,$$

where

$$\begin{aligned} \bar{G}_n(\vartheta, h)(y, s) &= \frac{1}{n} \sum_{i=1}^n I\{\Lambda_\vartheta(Y_i) - h(X_i) \leq y\} I\{X_i \leq s\} \\ &\quad - F_X(s) \frac{1}{n} \sum_{i=1}^n I\{\Lambda_\vartheta(Y_i) - h(X_i) \leq y\}. \end{aligned} \quad (\text{A.7})$$

Note that for any deterministic function  $h$  we have  $\mathbb{E}[\bar{G}_n(\vartheta, h)] = G(\vartheta, h)$ . The assertion of the theorem follows from

$$\sup_{\vartheta \in \Theta} \|G_n(\vartheta, \hat{h}_\vartheta) - \bar{G}_n(\vartheta, \hat{h}_\vartheta)\| \leq \sup_{s \in [0,1]} |\hat{F}_{X,n}(s) - F_X(s)| = o_P(1)$$

and Lemmas A.2 and A.3.  $\square$

**Lemma A.2** *Under the assumptions of Theorem 4.1 (i),*

$$\sup_{\vartheta \in \Theta} \|\bar{G}_n(\vartheta, \hat{h}_\vartheta) - G(\vartheta, \hat{h}_\vartheta)\| = o_P(1).$$

**Proof.** From Lemma 4.2 follows the existence of some deterministic sequence  $a_n \searrow 0$  such that the probability of the event

$$\sup_{\vartheta \in \Theta} \|h_\vartheta - \hat{h}_\vartheta\|_\infty \leq a_n \quad (\text{A.8})$$

converges to one. Thus we assume in what follows that (A.8) holds.

We only consider the difference between the first sum in the definition of  $\bar{G}_n(\vartheta, h)$  (see (A.7)) and the first integral in the definition of  $G(\vartheta, h)$  (see (3.5)). The difference between the second sum and the second integral can be treated similarly. Applying (A.8) the first sum in  $\bar{G}_n(\vartheta, \hat{h}_\vartheta)(y, s)$  can be nested as

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n I\{\Lambda_\vartheta(Y_i) - h_\vartheta(X_i) \leq y - a_n\} I\{X_i \leq s\} \\ &\leq \frac{1}{n} \sum_{i=1}^n I\{\Lambda_\vartheta(Y_i) - \hat{h}_\vartheta(X_i) \leq y\} I\{X_i \leq s\} \\ &\leq \frac{1}{n} \sum_{i=1}^n I\{\Lambda_\vartheta(Y_i) - h_\vartheta(X_i) \leq y + a_n\} I\{X_i \leq s\} \end{aligned}$$

while the first integral in  $G(\vartheta, \hat{h}_\vartheta)(y, s)$  can be nested as

$$\int F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y - a_n + h_\vartheta(x))) - h_0(x)) I\{x \leq s\} f_X(x) dx$$

$$\begin{aligned}
&\leq \int F_0 \left( \Lambda_0(\Lambda_\vartheta^{-1}(y + \hat{h}_\vartheta(x))) - h_0(x) \right) I\{x \leq s\} f_X(x) dx \\
&\leq \int F_0 \left( \Lambda_0(\Lambda_\vartheta^{-1}(y + a_n + h_\vartheta(x))) - h_0(x) \right) I\{x \leq s\} f_X(x) dx.
\end{aligned}$$

Thus we have to consider

$$\begin{aligned}
&H_{n,\vartheta}^{(1)}(y, s) \\
&= \frac{1}{n} \sum_{i=1}^n \left( I\{\Lambda_\vartheta(Y_i) - h_\vartheta(X_i) \leq y + a_n\} I\{X_i \leq s\} \right. \\
&\quad \left. - \int F_0 \left( \Lambda_0(\Lambda_\vartheta^{-1}(y + a_n + h_\vartheta(x))) - h_0(x) \right) I\{x \leq s\} f_X(x) dx \right) \\
&H_{n,\vartheta}^{(2)}(y, s) \\
&= \int \left( F_0 \left( \Lambda_0(\Lambda_\vartheta^{-1}(y + a_n + h_\vartheta(x))) - h_0(x) \right) \right. \\
&\quad \left. - F_0 \left( \Lambda_0(\Lambda_\vartheta^{-1}(y + h_\vartheta(x))) - h_0(x) \right) \right) I\{x \leq s\} f_X(x) dx
\end{aligned}$$

and the same terms with  $y + a_n$  replaced by  $y - a_n$ , which can be treated completely analogously. We have to show that  $\sup_{\vartheta \in \Theta} \|H_{n,\vartheta}^{(1)}\| = o_P(1)$  and  $\sup_{\vartheta \in \Theta} \|H_{n,\vartheta}^{(2)}\| = o(1)$ .

Recall condition **(N1)** and note that  $\sup_{\vartheta \in \Theta} \sup_{\substack{s \in [0,1] \\ y \in C}} |H_{n,\vartheta}^{(2)}(y, s)| = o(1)$  follows from uniform continuity of  $F_0$  and of  $\Lambda_0 \circ \Lambda_\vartheta^{-1}$  uniformly in  $\vartheta$  (see **(B5)** and **(B4)**), from the representation  $h_\vartheta = \Lambda_\vartheta \circ \Lambda_0^{-1} \circ h_0$  and uniform continuity of  $\Lambda_\vartheta$  uniformly in  $\vartheta$  (see **(B3)**), and  $a_n \rightarrow 0$ .

Let  $n$  be large enough such that  $|a_n| \leq \tau$  for  $\tau$  both from **(B5)** and **(B4)**. Now to prove  $\sup_{\vartheta \in \Theta} \|H_{n,\vartheta}^{(1)}\| = o_P(1)$  note that

$$\sup_{\vartheta \in \Theta} \|H_{n,\vartheta}^{(1)}\| \leq \sup_{f \in \mathcal{F}} |P_n f - P f|,$$

where  $P_n$  denotes the empirical measure of  $(X_1, Y_1), \dots, (X_n, Y_n)$ , and  $P$  the measure of  $(X_1, Y_1)$ , and

$$\mathcal{F} = \{(x, y) \mapsto I\{\Lambda_\vartheta(y) - h_\vartheta(x) \leq z\} I\{x \leq s\} \mid \vartheta \in \Theta, s \in [0, 1], z \in C_\tau\}$$

with  $C_\tau$  as in assumption **(B4)**. The assertion follows from the Glivenko-Cantelli theorem as stated in Theorem 2.4.1 in van der Vaart and Wellner (1996) if we show that the bracketing number  $N_{[]}(\epsilon, \mathcal{F}, L_1(P))$  is finite for each  $\epsilon > 0$ . To this end let  $\epsilon > 0$  and for the moment fix  $s \in [0, 1]$ ,  $\vartheta \in \Theta$  and  $z \in C_\tau$ . Choose  $\delta > 0$  corresponding to  $\epsilon$  as in assumption **(B5)**.

Partition  $[0, 1]$  into finitely many intervals  $[s_j, s_{j+1}]$  such that  $F_X(s_{j+1}) - F_X(s_j) \leq \epsilon$  for all  $j$ . For the fixed  $s$ , denote the interval containing  $s$  by  $[s_j, s_{j+1}] = [s^\ell, s^u]$ .

Now choose a finite sup-norm bracketing of length  $\gamma$  for the class  $\mathcal{L}_S = \{\Lambda_\vartheta|_S : \vartheta \in \Theta\}$  according to (4.1) with  $\gamma$  as in assumption **(B4)** corresponding to the above chosen  $\delta$ . For the fixed  $\vartheta$  this gives a bracket  $h^\ell \leq h_\vartheta \leq h^u$  of sup-norm length  $\gamma$ .

Choose a finite sup-norm bracketing of length  $\delta$  for the class  $\mathcal{L}_S^1 = \{\Lambda_0 \circ \Lambda_\vartheta^{-1}|_S : \vartheta \in \Theta\}$  according to (4.1). For the fixed  $\vartheta$  this gives a bracket  $V^\ell \leq \Lambda_0 \circ \Lambda_\vartheta^{-1} \leq V^u$ .

Then consider the bounded and increasing function

$$D(z) = \int F_0(V^\ell(z + h^\ell(x)) - h_0(x)) f_X(x) dx$$

and choose a finite partition of the compact  $C_\tau$  in intervals  $[z_k, z_{k+1}]$  such that  $D(z_{k+1}) - D(z_k) < \epsilon$ . For the fixed  $z$ , denote the interval containing  $z$  by  $[z_k, z_{k+1}] = [z^\ell, z^u]$ .

Now for the function  $f \in \mathcal{F}$  that is determined by  $\vartheta$ ,  $s$  and  $z$ , a bracket is given by  $[f^\ell, f^u]$  with

$$\begin{aligned} f^\ell(x, y) &= I\{\Lambda_0(y) \leq V^\ell(z^\ell + h^\ell(x))\} I\{x \leq s^\ell\} \\ f^u(x, y) &= I\{\Lambda_0(y) \leq V^u(z^u + h^u(x))\} I\{x \leq s^u\} \end{aligned}$$

with  $L_1(P)$ -norm

$$\begin{aligned} & \mathbb{E}[I\{\Lambda_0(Y_i) \leq V^u(z^u + h^u(X_i))\} I\{X_i \leq s^u\}] - \mathbb{E}[I\{\Lambda_0(Y_i) \leq V^\ell(z^\ell + h^\ell(X_i))\} I\{X_i \leq s^\ell\}] \\ & \leq F_X(s^u) - F_X(s^\ell) \\ & \quad + \int \left| F_0(V^u(z^u + h^u(x)) - h_0(x)) - F_0(V^\ell(z^\ell + h^\ell(x)) - h_0(x)) \right| f_X(x) dx \\ & \leq 2\epsilon + \int \left| F_0(V^u(z^u + h^u(x)) - h_0(x)) - F_0(\Lambda_0(\Lambda_\vartheta^{-1}(z^u + h^u(x))) - h_0(x)) \right| f_X(x) dx \\ & \quad + \int \left| F_0(\Lambda_0(\Lambda_\vartheta^{-1}(z^u + h^\ell(x))) - h_0(x)) - F_0(V^\ell(z^u + h^\ell(x)) - h_0(x)) \right| f_X(x) dx \\ & \quad + \int \left| F_0(\Lambda_0(\Lambda_\vartheta^{-1}(z^u + h^u(x))) - h_0(x)) - F_0(\Lambda_0(\Lambda_\vartheta^{-1}(z^\ell + h^\ell(x))) - h_0(x)) \right| f_X(x) dx \\ & \leq 4\epsilon \end{aligned}$$

by the definition of  $[s^\ell, s^u]$  and  $[z^\ell, z^u]$  and using the construction of brackets above (note that  $\|V^u - \Lambda_0 \circ \Lambda_\vartheta^{-1}\|_\infty \leq \delta$ ,  $\|\Lambda_0 \circ \Lambda_\vartheta^{-1} - V^\ell\|_\infty \leq \delta$ ,  $\|h^u - h^\ell\|_\infty \leq \gamma$  and recall assumptions **(B5)** and **(B4)**).

There are finitely many such brackets to cover  $\mathcal{F}$  and thus the assertion follows.  $\square$

**Lemma A.3** *Under the assumptions of Theorem 4.1 (i),*

$$\sup_{\vartheta \in \Theta} \|G(\vartheta, h_\vartheta) - G(\vartheta, \hat{h}_\vartheta)\| = o_P(1).$$

**Proof.** According to assumption **(N1)** it suffices to show

$$\sup_{\vartheta \in \Theta} \sup_{\substack{s \in [0,1] \\ y \in C}} |G(\vartheta, h_\vartheta)(y, s) - G(\vartheta, \hat{h}_\vartheta)(y, s)| = o_P(1).$$

Recalling the definition of  $G$  in (3.5) we see that the assertion follows from Lemma 4.2 and uniform continuity of  $F_0$  and of  $\Lambda_0 \circ \Lambda_\vartheta^{-1}$  (uniformly in  $\vartheta$ ).  $\square$

## B Identifiability of the model in the random design case

We prove the assertion of Remark 2.2. First note that  $\varepsilon(\vartheta_1)$  is independent of  $X$ , and thus the conditional distribution of  $\varepsilon(\vartheta_1)$ , i.e.

$$\begin{aligned} \mathbb{P}(\varepsilon(\vartheta_1) \leq y \mid X = x) &= \mathbb{P}(Y \leq \Lambda_{\vartheta_1}^{-1}(y + h_{\vartheta_1}(x)) \mid X = x) \\ &= F_0(\Lambda_{\vartheta_0}(\Lambda_{\vartheta_1}^{-1}(y + h_{\vartheta_1}(x))) - h_{\vartheta_0}(x)) \end{aligned}$$

does not depend on  $x$ . Further,  $h_{\vartheta_0} = \Lambda_{\vartheta_0} \circ \Lambda_{\vartheta_1}^{-1} \circ h_{\vartheta_1}$ , and for  $y \leq 0$  we have  $\Lambda_{\vartheta_0}(\Lambda_{\vartheta_1}^{-1}(y + h_{\vartheta_1}(x))) \leq \Lambda_{\vartheta_0}(\Lambda_{\vartheta_1}^{-1}(h_{\vartheta_1}(x)))$  because  $\Lambda_{\vartheta_0} \circ \Lambda_{\vartheta_1}^{-1}$  is strictly increasing. As  $F_0$  is strictly increasing by assumption it follows that

$$H^{-1}(y + H(h_{\vartheta_0}(x))) - h_{\vartheta_0}(x)$$

does not depend on  $x$  for  $y \in (-\infty, 0]$  and  $x \in [0, 1]$ , where for ease of presentation write  $H := \Lambda_{\vartheta_1} \circ \Lambda_{\vartheta_0}^{-1}$ . Thus

$$H^{-1}(y + H(a)) - a = H^{-1}(y + H(b)) - b$$

for all  $y \leq 0$ ,  $a, b \in h_{\vartheta_0}([0, 1])$ . Because  $Y$  may take the value 0 by assumption and  $\varepsilon \leq 0$  one obtains  $h_{\vartheta_0}([0, 1]) \cap \mathbb{R}_0^+ \neq \emptyset$ . To conclude the proof we distinguish two cases.

(1) Let  $h_{\vartheta_0}([0, 1]) \cap \mathbb{R}_0^+ = \{0\}$ . Set  $a = 0$ , since by assumption  $\Lambda_\vartheta(0) = 0$  for all  $\vartheta \in \Theta$ , then

$$H^{-1}(y) = H^{-1}(y + H(b)) - b$$

for all  $y \leq 0$ ,  $b \in h_{\vartheta_0}([0, 1]) \subset \mathbb{R}_0^-$ . Set  $c = H^{-1}(y + H(b))$ , then it follows that

$$H(c) - H(b) = H(c - b)$$

for all  $b, c \in (-\delta, 0]$  for some  $\delta > 0$  and from the assumptions it follows that  $\vartheta_1 = \vartheta_0$  with  $H = \text{id}$ .

(2) Let  $h_{\vartheta_0}([0, 1]) \cap \mathbb{R}_0^+ = I$  be an interval of positive length. For  $a \in I$  one has  $y := -H(a) \leq 0$ , and

$$0 = H^{-1}(0) = a + H^{-1}(-H(a) + H(b)) - b$$

and thus

$$H(b - a) = H(b) - H(a)$$

for all  $a, b \in I$ . From the assumptions it follows that  $\vartheta_1 = \vartheta_0$  with  $H = \text{id}$  and thus identifiability of the model. □

## References

- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *J. Roy. Statist. Soc. Ser. 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.
- Carroll, R. J. and Ruppert, D. (1988). *Transformation and weighting in regression*. Monographs on Statistics and Applied Probability. Chapman & Hall, New York.
- Colling, B. and Van Keilegom, I. (2016). Goodness-of-fit tests in semiparametric transformation models. *TEST* **25**, 291–308.
- Daouia, A., Noh, H. and Park, B. U. (2016). Data envelope fitting with constrained polynomial splines. *J. R. Stat. Soc. B.* **78**, 3–30.
- Drees, H., Neumeier, N. and Selk, L. (2018). Estimation and hypotheses tests in boundary regression models. *Bernoulli*, to appear.  
<http://www.bernoulli-society.org/index.php/publications/bernoulli-journal>
- Girard, S. and Jacob, P. (2008). Frontier estimation via kernel regression on high power-transformed data. *J. Multivariate Anal.* **99**, 403–420.
- Hall, P. and Van Keilegom, I. (2009). Nonparametric "regression" when errors are positioned at end-points. *Bernoulli* **15**, 614–633.
- Hall, P., Park, B.U. and Stern, S.E. (1998). On polynomial estimators of frontiers and boundaries. *J. Multivariate Anal.* **66**, 71–98.
- Härdle, W., Park, B.U. and Tsybakov, A.B. (1995). Estimation of a non sharp support boundaries. *J. Multivariate Anal.* **55**, 205–218.

- Horowitz, J. L. (2009). *Semiparametric and nonparametric methods in econometrics*. Springer Series in Statistics. Springer, New York.
- Jirak, M., Meister, A. and Reiß, M. (2013). Asymptotic equivalence for nonparametric regression with non-regular errors. *Probability theory and relative fields* **155**, 201–229.
- Jirak, M., Meister, A. and Reiß, M. (2014). Adaptive estimation in nonparametric regression with one-sided errors. *Ann. Statist.* **42**, 1970–2002.
- Jones, M.C. and Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika* **96**, 761–780.
- Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York.
- Linton, O., Sperlich, S. and Van Keilegom, I. (2008). Estimation on a semiparametric transformation model. *Ann. Statist.* **36**, 686–718.
- Manly, B.F.J (1976). Exponential data transformations. *Journal of the Royal Statistical Society. Series D* **25**, 37–42.
- Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance* **7**, 77–91.
- Meister, A. and Reiß, M. (2013). Asymptotic equivalence for nonparametric regression with non-regular errors. *Probab. Th. Rel. Fields* **155**, 201–229.
- Mu, Y. and He, X. (2007). Power transformation toward a linear regression quantile. *J. Amer. Statist. Assoc.* **102**, 269–279.
- Müller, U.U. and Wefelmeyer W. (2010). Estimation in nonparametric regression with non-regular errors. *Comm. Statist. Theory Methods* **39**, 1619–1629.
- Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.
- Powell, J. (1991). Estimation of monotonic regression models under quantile restrictions. In: Barnett, W., Powell, J., and Tauchen, G. (eds.), *Nonparametric and Semiparametric Methods in Econometrics*, 357–384. Cambridge University Press, New York.
- Simar, L. and Wilson, P.W. (1998). Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Management Science* **44**, 49–61.
- van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes*. Springer, New York.
- Wilson, P.W. (2003). Testing independence in models of productive efficiency. *Journal of Productivity Analysis* **20**, 361–390.

Yeo, I-K. and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika* **87**, 954–959.

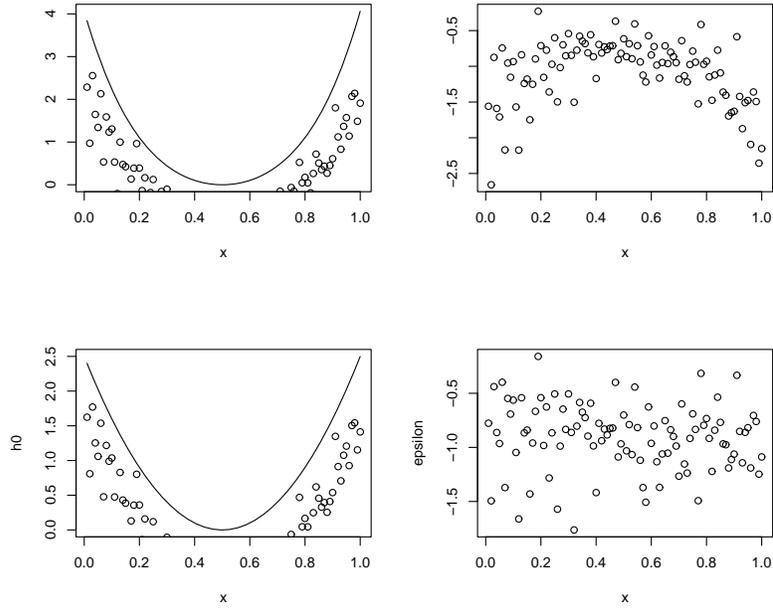


Figure 1: Original data (upper panel) and transformed data (2.2) (lower panel) with  $h_0(x) = 10(x - \frac{1}{2})^2$ ,  $n = 100$  and  $\varepsilon_{i,n} \sim \text{Weibull}(1, 3)$  with Yeo and Johnson transformation  $\Lambda_{0.5}$  as defined in Example 2.3. The design points are equidistant. The left panels show the data and regression functions, the right panels show the errors.

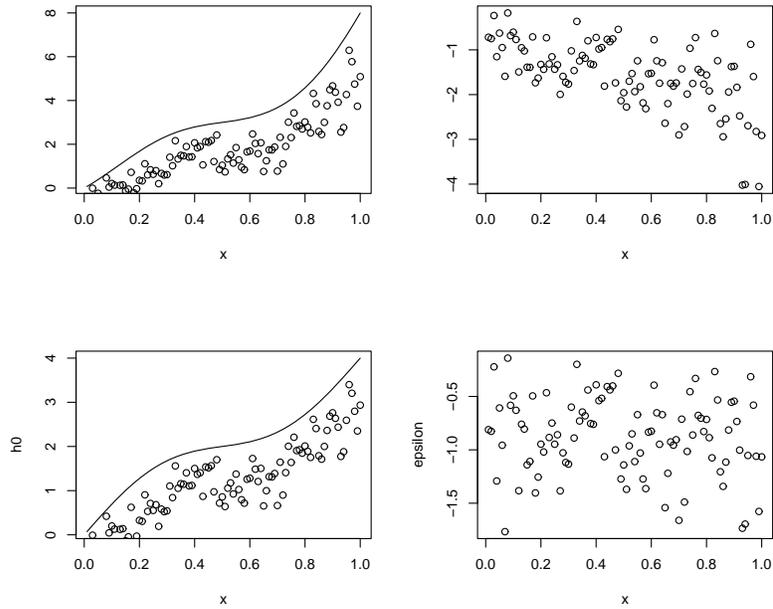


Figure 2: The setting is similar to Figure 1 but with  $h_0(x) = \frac{1}{2} \sin(2\pi x) + 4x$ .

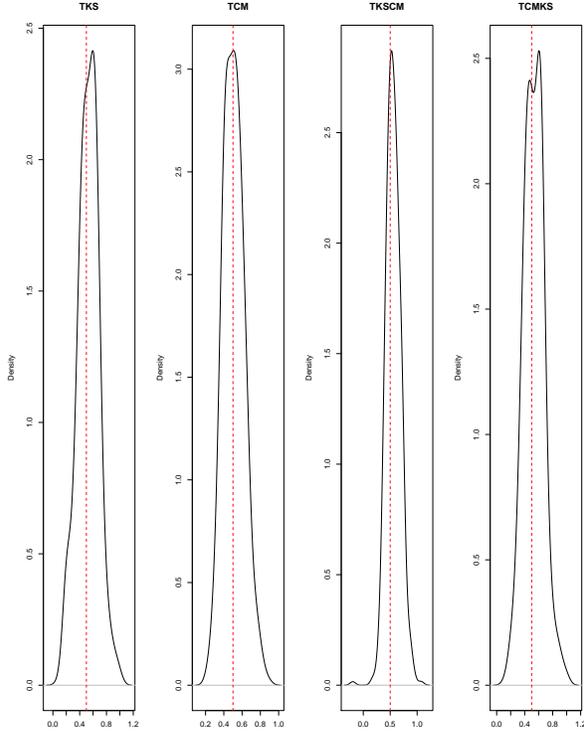


Figure 3: Density function of the four estimators TKS, TCM, TCMKS and TKSCM for the Model (5.1) with a sample size  $n = 100$  and bandwidths  $a_n = b_n/20$  with  $b_n = n^{-1/3}$ . This corresponds to results in Table 2. The vertical dashed line corresponds the true parameter  $\vartheta_0 = 0.5$ .

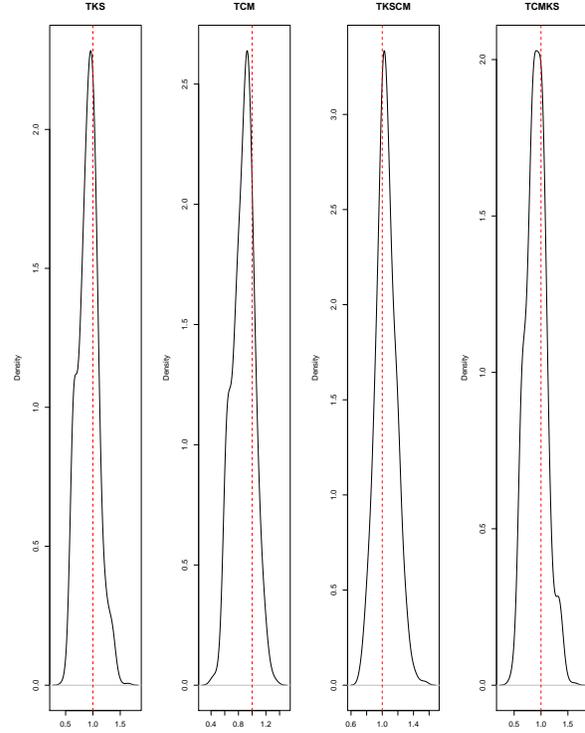


Figure 4: Density function of the four estimators TKS, TCM, TCMKS and TKSCM for the Model (5.4) with a sample size  $n = 100$  and bandwidths  $a_n = b_n/2$  with  $b_n = n^{-1/3}$ . This corresponds to results in Table 3. The vertical dashed line corresponds the true parameter  $\vartheta_0 = 1$ .

$n = 50$	TKS	TCM	TKSCM	TCMKS
$\vartheta_0 = 0$	0.162 0.162 (0.120)	<b>0.095</b> 0.122 ( <u>0.060</u> )	0.216 0.241 (0.149)	0.196 0.183 (0.141)
$\vartheta_0 = 0.5$	0.643 0.646 (0.142)	<b>0.576</b> 0.595 ( <u>0.080</u> )	0.766 0.800 (0.250)	0.691 0.646 (0.204)
$\vartheta_0 = 1$	1.120 1.190 (0.232)	<b>1.100</b> 1.090 ( <u>0.121</u> )	1.340 1.380 (0.335)	1.200 1.310 (0.287)
$\vartheta_0 = 1.5$	<b>1.610</b> 1.720 (0.228)	1.640 1.670 ( <u>0.133</u> )	1.900 1.920 (0.336)	1.620 1.780 (0.282)
$\vartheta_0 = 2$	1.810 2.020 (0.357)	<b>2.110</b> 2.130 ( <u>0.076</u> )	2.310 2.460 (0.179)	1.860 2.060 (0.297)

$n = 100$	TKS	TCM	TKSCM	TCMKS
$\vartheta_0 = 0$	0.014 -0.039 (0.055)	-0.014 -0.029 ( <u>0.019</u> )	0.098 0.092 (0.031)	<b>-0.006</b> -0.021 (0.026)
$\vartheta_0 = 0.5$	0.483 0.496 (0.041)	0.461 0.451 ( <u>0.026</u> )	0.625 0.618 (0.049)	<b>0.503</b> 0.523 (0.047)
$\vartheta_0 = 1$	0.951 0.964 (0.062)	0.949 0.950 ( <u>0.034</u> )	1.150 1.140 (0.059)	<b>1.000</b> 1.010 (0.071)
$\vartheta_0 = 1.5$	<b>1.500</b> 1.490 (0.050)	1.470 1.450 ( <u>0.040</u> )	1.670 1.660 (0.078)	1.520 1.520 (0.077)
$\vartheta_0 = 2$	1.960 2.000 (0.071)	<b>1.970</b> 1.960 ( <u>0.034</u> )	2.150 2.150 (0.065)	1.970 2.030 (0.074)

Table 1: Mean, median and MISE for Model (5.1) for  $n = 50$  and  $n = 100$  with  $a_n = b_n/2$ .

$n = 50$	TKS	TCM	TKSCM	TCMKS
$\vartheta_0 = 0$	0.154 0.146 (0.126)	<b>-0.059</b> -0.051 ( <u>0.050</u> )	0.198 0.205 (0.114)	0.173 0.171 (0.128)
$\vartheta_0 = 0.5$	0.613 0.635 (0.118)	<b>0.511</b> 0.513 ( <u>0.066</u> )	0.717 0.735 (0.173)	0.658 0.646 (0.163)
$\vartheta_0 = 1$	1.120 1.190 (0.182)	<b>1.030</b> 1.030 ( <u>0.091</u> )	1.280 1.300 (0.230)	1.190 1.270 (0.221)
$\vartheta_0 = 1.5$	1.630 1.690 (0.164)	<b>1.560</b> 1.540 ( <u>0.095</u> )	1.830 1.810 (0.233)	1.620 1.740 (0.209)
$\vartheta_0 = 2$	1.880 2.040 (0.272)	<b>2.050</b> 2.080 ( <u>0.066</u> )	2.280 2.370 (0.149)	1.870 2.040 (0.272)

$n = 100$	TKS	TCM	TKSCM	TCMKS
$\vartheta_0 = 0$	<b>0.003</b> 0.046 (0.040)	0.045 0.057 ( <u>0.020</u> )	0.078 0.067 (0.026)	0.005 0.022 (0.025)
$\vartheta_0 = 0.5$	0.454 0.448 (0.037)	0.418 0.406 ( <u>0.028</u> )	0.581 0.566 (0.033)	<b>0.473</b> 0.478 (0.046)
$\vartheta_0 = 1$	0.938 0.950 (0.053)	0.913 0.918 ( <u>0.038</u> )	1.100 1.080 (0.047)	<b>0.984</b> 0.986 (0.070)
$\vartheta_0 = 1.5$	1.450 1.420 (0.052)	1.410 1.390 ( <u>0.041</u> )	1.600 1.590 (0.050)	<b>1.470</b> 1.440 (0.062)
$\vartheta_0 = 2$	<b>1.950</b> 1.960 (0.057)	1.930 1.910 ( <u>0.037</u> )	2.100 2.100 (0.053)	1.940 1.980 (0.066)

Table 2: Mean, median and MISE for Model (5.1) for  $n = 50$  and  $n = 100$  with  $a_n = b_n/20$ .

$n = 50$	TKS	TCM	TKSCM	TCMKS
$\vartheta_0 = 0$	0.054 0.062 (0.218)	-0.029 -0.029 (0.005)	<b>0.000</b> 0.000 ( <u>0.000</u> )	0.050 0.046 (0.182)
$\vartheta_0 = 0.5$	0.419 0.427 (0.038)	0.412 0.408 (0.027)	<b>0.469</b> 0.478 ( <u>0.025</u> )	0.458 0.470 (0.030)
$\vartheta_0 = 1$	0.887 0.881 (0.080)	0.791 0.770 (0.104)	<b>0.942</b> 0.971 (0.077)	0.934 0.933 ( <u>0.069</u> )
$\vartheta_0 = 1.5$	1.360 1.350 (0.130)	1.210 1.270 (0.214)	1.310 1.420 (0.319)	<b>1.410</b> 1.350 ( <u>0.111</u> )
$\vartheta_0 = 2$	1.800 1.790 (0.165)	1.560 1.690 (0.504)	1.249 1.769 (1.739)	<b>1.810</b> 1.790 ( <u>0.141</u> )

$n = 100$	TKS	TCM	TKSCM	TCMKS
$\vartheta_0 = 0$	0.025 0.043 (0.004)	-0.037 -0.038 ( <u>0.003</u> )	<b>0.012</b> 0.009 ( <u>0.002</u> )	-0.034 -0.054 (0.004)
$\vartheta_0 = 0.5$	0.459 0.458 (0.017)	0.406 0.405 (0.016)	<b>0.520</b> 0.518 ( <u>0.008</u> )	0.449 0.453 (0.017)
$\vartheta_0 = 1$	0.922 0.945 (0.038)	0.845 0.856 (0.046)	<b>1.030</b> 1.040 ( <u>0.017</u> )	0.907 0.909 (0.044)
$\vartheta_0 = 1.5$	1.430 1.350 (0.056)	1.290 1.310 (0.080)	<b>1.540</b> 1.540 ( <u>0.032</u> )	1.400 1.350 (0.063)
$\vartheta_0 = 2$	<b>1.930</b> 1.970 ( <u>0.072</u> )	1.740 1.770 (0.126)	1.880 2.010 (0.419)	1.850 1.790 (0.090)

Table 3: Mean, median and MISE for Model (5.4) for  $n = 50$  and  $n = 100$  with  $a_n = b_n/2$ .

$n = 50$	TKS	TCM	TKSCM	TCMKS
$\vartheta_0 = 0$	0.021 0.062 (0.177)	0.053 0.057 ( <u>0.007</u> )	<b>0.013</b> 0.010 ( <u>0.007</u> )	0.017 0.062 (0.141)
$\vartheta_0 = 0.5$	0.404 0.415 (0.044)	0.375 0.380 (0.034)	<b>0.449</b> 0.461 ( <u>0.029</u> )	0.434 0.445 (0.034)
$\vartheta_0 = 1$	0.851 0.830 (0.086)	0.737 0.695 (0.119)	<b>0.915</b> 0.931 ( <u>0.074</u> )	0.895 0.891 (0.076)
$\vartheta_0 = 1.5$	1.330 1.350 (0.150)	1.130 1.200 (0.260)	1.310 1.410 (0.293)	<b>1.380</b> 1.350 ( <u>0.119</u> )
$\vartheta_0 = 2$	1.760 1.790 (0.176)	1.520 1.610 (0.435)	1.409 1.770 (1.250)	<b>1.770</b> 1.790 ( <u>0.156</u> )

$n = 100$	TKS	TCM	TKSCM	TCMKS
$\vartheta_0 = 0$	0.027 0.043 (0.004)	-0.038 -0.040 ( <u>0.003</u> )	<b>0.011</b> 0.008 ( <u>0.002</u> )	0.037 0.062 (0.004)
$\vartheta_0 = 0.5$	0.462 0.466 (0.018)	0.411 0.407 (0.015)	<b>0.524</b> 0.522 ( <u>0.008</u> )	0.453 0.460 (0.019)
$\vartheta_0 = 1$	0.930 0.950 (0.037)	0.849 0.856 (0.043)	<b>1.040</b> 1.030 ( <u>0.018</u> )	0.917 0.915 (0.042)
$\vartheta_0 = 1.5$	1.440 1.390 (0.058)	1.300 1.320 (0.078)	<b>1.550</b> 1.540 ( <u>0.031</u> )	1.420 1.350 (0.065)
$\vartheta_0 = 2$	<b>1.930</b> 1.960 ( <u>0.067</u> )	1.730 1.760 (0.130)	1.830 2.010 (0.509)	1.840 1.790 (0.091)

Table 4: Mean, median and MISE for Model (5.4) for  $n = 50$  and  $n = 100$  with  $a_n = b_n/20$ .

Method	Pearson	Kendall	Spearman
Original data	-0.273	-0.165	-0.234
True parameter $\vartheta_0$	0.005	0.003	0.004
TKS	0.008	0.004	0.007
TCM	0.024	0.014	0.021
TKSCM	0.011	0.007	0.009
TCMKS	0.003	0.001	0.002

Table 5: *Pearson's, Kendall's and Spearman's correlation coefficients (the average over 1000 iterations) between the covariates and the errors for the model (5.4) when  $n = 100$ . The first line corresponds to the correlations for the original data while the second line is for the true transformation parameter ( $\vartheta_0 = 0.5$ ). The last four lines correspond to the correlations for each estimator.*

# Supplementary material to “Semi-parametric transformation boundary regression models”

## C Proofs of asymptotic results in the fixed design case

### C.1 Proof of Lemma 3.3

To prove Lemma 3.3, we first need the following technical lemma.

**Lemma C.1** *Assume model (2.2) holds under assumptions (A1’), (A2’) and (A4’). Then we have*

$$\sup_{x \in [0,1]} \min_{\substack{i \in \{1, \dots, n\} \\ |x_{i,n} - x| \leq b_n}} |\varepsilon_{i,n}| = o_P(1).$$

**Proof.** The proof is similar to the proof of Lemma A.2 in Drees et al. (2018) but some adaptations are needed to deal with non-equidistant fixed design points. Let  $Z_1, Z_2, \dots$  be iid with the same distribution as  $-\varepsilon_{i,n}$  with cumulative distribution function  $U$ . To prove the result, we shall show that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \sup_{x \in [0,1]} \min_{\substack{i \in \{1, \dots, n\} \\ |x_{i,n} - x| \leq b_n}} Z_i > \epsilon \right) = 0, \quad \epsilon > 0.$$

For  $n \geq 1$ , let  $0 < k \leq n$ ,  $x \in [0, 1]$  and set  $I_n = [x - b_n, x + b_n]$ . Assume that exactly  $k$  points lie in  $I_n$ , say

$$x_{m+1,n} < \dots < x_{m+k,n} \in I_n$$

for some  $m < n + 1 - k$ . We shall distinguish two cases.

(1) If  $(x_{m,n}, x_{m+k+1,n}) \in [0, 1]^2$ , it means that

$$2b_n = |I_n| < x_{m+k+1,n} - x_{m,n} = \sum_{j=m}^{m+k} (x_{j+1,n} - x_{j,n}) \leq (k+1)\bar{\Delta}_n,$$

since  $\bar{\Delta}_n \geq x_{j,n} - x_{j-1,n}$  for any  $1 \leq j \leq n+1$ .

(2) If  $x_{m,n}$  or  $x_{m+k+1,n}$  do not exist, which means that either  $x_{m+1,n} = x_{0,n} = 0$  or  $x_{m+k+1,n} = x_{n+1,n} = 1$ . Consider the first case  $x_{m+1,n} = x_{0,n}$  (the extremal case is  $x = 0$ ). Then we have

$$b_n = \frac{|I_n|}{2} < x_{k,n} - x_{0,n} = \sum_{j=0}^{k-1} (x_{j+1,n} - x_{j,n}) \leq k\bar{\Delta}_n.$$

A similar inequality holds for  $x_{m+k+1,n} = x_{n+1,n} = 1$  (with the extremal case  $x = 1$ ).

In both cases, (1) and (2) yield to

$$b_n < k\bar{\Delta}_n \Rightarrow k > \frac{b_n}{\bar{\Delta}_n}, \quad n \geq 1.$$

Then, for all  $y > 0$ , we have with  $d_n := \lceil \frac{b_n}{\bar{\Delta}_n} \rceil$

$$\begin{aligned} \mathbb{P} \left( \sup_{x \in [0,1]} \min_{\substack{i \in \{1, \dots, n\} \\ |x_{i,n} - x| \leq b_n}} Z_i > y \right) &\leq \mathbb{P} \left( \left\{ \max_{j \in \{1, \dots, n-d_n\}} \min_{i \in \{j, \dots, j+d_n\}} Z_i > y \right\} \right) \\ &\leq \sum_{j=1}^{n-d_n} \mathbb{P} \left( \min_{i \in \{j, \dots, j+d_n\}} Z_i > y \right) \\ &= (n - d_n) \mathbb{P} \left( \min_{i \in \{1, \dots, d_n+1\}} Z_i > y \right) \\ &= (n - d_n) \bar{U}(y)^{d_n+1}. \end{aligned}$$

Thus it remains to show that for all  $\epsilon > 0$

$$(n - d_n) \bar{U}(\epsilon)^{d_n+1} \xrightarrow[n \rightarrow \infty]{} 0$$

which is true since  $d_n \underset{n \rightarrow \infty}{\sim} \frac{b_n}{\bar{\Delta}_n}$  and

$$\begin{aligned} \frac{b_n}{\bar{\Delta}_n} \log(\bar{U}(\epsilon)) + \log(n - d_n) &\leq \frac{b_n}{\bar{\Delta}_n} \log(\bar{U}(\epsilon)) + \log(n) \\ &= \log(n) \left( \frac{b_n}{\bar{\Delta}_n \log(n)} \log(\bar{U}(\epsilon)) + 1 \right) \\ &\xrightarrow[n \rightarrow \infty]{} -\infty \end{aligned}$$

since  $\bar{U}(\epsilon) < 1$  under (A1') and  $\frac{b_n}{\bar{\Delta}_n \log(n)} \xrightarrow[n \rightarrow \infty]{} \infty$  under (A4'). This concludes the proof.  $\square$

The **proof of Lemma 3.3** is analogous to the proof of Lemma 3.1.

## C.2 Proof of Theorem 4.1 in the fixed design case

The first part of the proof is similar to the random design case. Here, we use

$$\begin{aligned} \sup_{\vartheta \in \Theta} |M_n(\vartheta) - M(\vartheta)| &\leq \sup_{\vartheta \in \Theta} \|G_n(\vartheta, \hat{h}_\vartheta) - \bar{G}_n(\vartheta, \hat{h}_\vartheta)\| + \sup_{\vartheta \in \Theta} \|\bar{G}_n(\vartheta, \hat{h}_\vartheta) - \tilde{G}_n(\vartheta, \hat{h}_\vartheta)\| \\ &\quad + \sup_{\vartheta \in \Theta} \|\tilde{G}_n(\vartheta, \hat{h}_\vartheta) - G(\vartheta, \hat{h}_\vartheta)\| + \sup_{\vartheta \in \Theta} \|G(\vartheta, \hat{h}_\vartheta) - G(\vartheta, h_\vartheta)\|, \end{aligned}$$

where the definition for  $M$  and  $G$  is as in the random case, and

$$\bar{G}_n(\vartheta, h)(y, s) = \frac{1}{n} \sum_{i=1}^n I\{\Lambda_\vartheta(Y_{i,n}) - h(x_{i,n}) \leq y\} (I\{x_{i,n} \leq s\} - F_X(s)).$$

Further,

$$\begin{aligned} \tilde{G}_n(\vartheta, h)(y, s) &= \frac{1}{n} \sum_{i=1}^n F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y + h(x_{i,n}))) - h_0(x_{i,n})) I\{x_{i,n} \leq s\} \\ &\quad - F_X(s) \sum_{i=1}^n F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y + h(x_{i,n}))) - h_0(x_{i,n})) \end{aligned} \quad (\text{C.1})$$

is a Riemann-sum approximation of  $G(\vartheta, h)(y, s)$ . Note that for any deterministic function  $h$  we have  $\tilde{G}_n(\vartheta, h) = \mathbb{E}[\bar{G}_n(\vartheta, h)]$ . The assertion of the theorem follows from

$$\sup_{\vartheta \in \Theta} \|G_n(\vartheta, \hat{h}_\vartheta) - \bar{G}_n(\vartheta, \hat{h}_\vartheta)\| \leq \sup_{s \in [0,1]} |\hat{F}_{X,n}(s) - F_X(s)| = o(1) \quad (\text{C.2})$$

and from Lemmas C.2–C.4 by an application of the arg-max theorem. For (C.2) note that with assumption **(A2'')**

$$\sup_{s \in [0,1]} |\hat{F}_{X,n}(s) - F_X(s)| = \sup_{s \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^n I\{x_{i,n} \leq s\} - \int_0^s f_X(x) dx \right| \quad (\text{C.3})$$

$$\begin{aligned} &\leq \sup_{s \in [0,1]} \left| \sum_{i=1}^n \int_{x_{i-1,n}}^{x_{i,n}} f_X(x) dx I\{x_{i,n} \leq s\} - \int_0^s f_X(x) dx \right| + o(1) \\ &= \sup_{s \in [0,1]} \left| \int_{\max\{x_{i,n} | x_{i,n} \leq s\}}^s f_X(x) dx \right| + o(1) \\ &= \bar{\Delta}_n \sup_{x \in [0,1]} f_X(x) + o(1) = o(1). \end{aligned} \quad (\text{C.4})$$

□

**Lemma C.2** *Under the assumptions of Theorem 4.1 (ii),*

$$\sup_{\vartheta \in \Theta} \|\bar{G}_n(\vartheta, \hat{h}_\vartheta) - \tilde{G}_n(\vartheta, \hat{h}_\vartheta)\| = o_P(1).$$

**Proof.** As in the proof of Lemma A.2 we assume in what follows that (A.8) holds. We only consider the difference between the first sum in the definitions of  $G_n(\vartheta, h)$  and the first sum in  $\tilde{G}_n(\vartheta, h)$  (see (3.6) and (C.1), respectively). The difference of the second sums can be treated similarly. Applying (A.8) the first sum in  $G_n(\vartheta, \hat{h}_\vartheta)(y, s)$  can be nested as

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n I\{\Lambda_\vartheta(Y_{i,n}) - h_\vartheta(x_{i,n}) \leq y - a_n\} I\{x_{i,n} \leq s\} \\ & \leq \frac{1}{n} \sum_{i=1}^n I\{\Lambda_\vartheta(Y_{i,n}) - \hat{h}_\vartheta(x_{i,n}) \leq y\} I\{x_{i,n} \leq s\} \\ & \leq \frac{1}{n} \sum_{i=1}^n I\{\Lambda_\vartheta(Y_{i,n}) - h_\vartheta(x_{i,n}) \leq y + a_n\} I\{x_{i,n} \leq s\} \end{aligned}$$

while the first sum in  $\tilde{G}_n(\vartheta, \hat{h}_\vartheta)(y, s)$  can be nested as

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y - a_n + h_\vartheta(x_{i,n}))) - h_0(x_{i,n})) I\{x_{i,n} \leq s\} \\ & \leq \frac{1}{n} \sum_{i=1}^n F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y + \hat{h}_\vartheta(x_{i,n}))) - h_0(x_{i,n})) I\{x_{i,n} \leq s\} \\ & \leq \frac{1}{n} \sum_{i=1}^n F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y + a_n + h_\vartheta(x_{i,n}))) - h_0(x_{i,n})) I\{x_{i,n} \leq s\}. \end{aligned}$$

Thus we have to consider

$$\begin{aligned} H_{n,\vartheta}^{(1)}(y, s) &= \frac{1}{n} \sum_{i=1}^n \left( I\{\Lambda_\vartheta(Y_{i,n}) - h_\vartheta(x_{i,n}) \leq y + a_n\} \right. \\ & \quad \left. - F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y + a_n + h_\vartheta(x_{i,n}))) - h_0(x_{i,n})) \right) I\{x_{i,n} \leq s\} \\ H_{n,\vartheta}^{(2)}(y, s) &= \frac{1}{n} \sum_{i=1}^n \left( F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y + a_n + h_\vartheta(x_{i,n}))) - h_0(x_{i,n})) \right. \\ & \quad \left. - F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y + h_\vartheta(x_{i,n}))) - h_0(x_{i,n})) \right) I\{x_{i,n} \leq s\} \end{aligned}$$

and the same terms with  $y + a_n$  replaced by  $y - a_n$ , which can be treated completely analogously. We have to show that  $\sup_{\vartheta \in \Theta} \|H_{n,\vartheta}^{(1)}\| = o_P(1)$  and  $\sup_{\vartheta \in \Theta} \|H_{n,\vartheta}^{(2)}\| = o(1)$ .

Recall condition **(N1)** and note that  $\sup_{\vartheta \in \Theta} \sup_{\substack{s \in [0,1] \\ y \in C}} |H_{n,\vartheta}^{(2)}(y, s)| = o(1)$  follows from uniform continuity of  $F_0$  and of  $\Lambda_0 \circ \Lambda_\vartheta^{-1}$  uniformly in  $\vartheta$  (see **(B5)** and **(B4)**), from the representation  $h_\vartheta = \Lambda_\vartheta \circ \Lambda_0^{-1} \circ h_0$  and uniform continuity of  $\Lambda_\vartheta$  uniformly in  $\vartheta$  (see **(B3)**), and  $a_n \rightarrow 0$ .

Now to prove  $\sup_{\vartheta \in \Theta} \|H_{n,\vartheta}^{(1)}\| = o_P(1)$ , let  $\epsilon > 0$  and for the moment fix  $s \in [0, 1]$ ,  $\vartheta \in \Theta$  and  $y \in C$ . Choose  $\delta > 0$  corresponding to  $\epsilon$  as in assumption **(B5)**. Let  $n$  be large enough such that  $|a_n| \leq \tau$  for  $\tau$  both from **(B5)** and **(B4)**.

Partition  $[0, 1]$  into finitely many intervals  $[s_j, s_{j+1}]$  such that  $F_X(s_{j+1}) - F_X(s_j) < \epsilon$  for all  $j$ . For the fixed  $s$ , denote the interval containing  $s$  by  $[s_j, s_{j+1}] = [s^\ell, s^u]$ .

Now choose a finite sup-norm bracketing of length  $\gamma$  for the class  $\mathcal{L}_S = \{\Lambda_\vartheta|_S : \vartheta \in \Theta\}$  according to (4.1) with  $\gamma$  as in assumption **(B4)** corresponding to the above chosen  $\delta$ . For the fixed  $\vartheta$  this gives a bracket  $h^\ell \leq h_\vartheta \leq h^u$  of sunorm length  $\gamma$ .

Choose a finite sup-norm bracketing of length  $\delta$  for the class  $\mathcal{L}_S^1 = \{\Lambda_0 \circ \Lambda_\vartheta^{-1}|_S : \vartheta \in \Theta\}$  according to (4.1). For the fixed  $\vartheta$  this gives a bracket  $V^\ell \leq \Lambda_0 \circ \Lambda_\vartheta^{-1} \leq V^u$ .

Then consider the bounded and increasing function

$$D_n(y) = \frac{1}{n} \sum_{i=1}^n F_0(V^\ell(y + a_n + h^\ell(x_{i,n})) - h_0(x_{i,n}))$$

and choose a finite partition of the compact  $C$  in intervals  $[y_k, y_{k+1}]$  such that  $D_n(y_{k+1}) - D_n(y_k) < \epsilon$ . For the fixed  $y$ , denote the interval containing  $y$  by  $[y_k, y_{k+1}] = [y^\ell, y^u]$ . Note that the brackets depend on  $n$ . This is suppressed in the notation because it is not relevant for the remainder of the proof because the number of brackets is  $O(\epsilon^{-1})$ , uniformly in  $n$ .

Now we can nest as follows

$$\begin{aligned} & I\{\Lambda_0(Y_{i,n}) \leq V^\ell(y^\ell + a_n + h^\ell(x_{i,n}))\} I\{x_{i,n} \leq s^\ell\} \\ & \leq I\{\Lambda_\vartheta(Y_{i,n}) - h_\vartheta(x_{i,n}) \leq y + a_n\} I\{x_{i,n} \leq s\} \\ & = I\{Y_{i,n} \leq \Lambda_\vartheta^{-1}(y + a_n + h_\vartheta(x_{i,n}))\} I\{x_{i,n} \leq s\} \\ & \leq I\{\Lambda_0(Y_{i,n}) \leq V^u(y^u + a_n + h^u(x_{i,n}))\} I\{x_{i,n} \leq s^u\}, \end{aligned}$$

and have

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \left( \mathbb{E}[I\{\Lambda_0(Y_{i,n}) \leq V^u(y^u + a_n + h^u(x_{i,n}))\} I\{x_{i,n} \leq s^u\}] \right. \\ & \quad \left. - \mathbb{E}[I\{\Lambda_0(Y_{i,n}) \leq V^\ell(y^\ell + a_n + h^\ell(x_{i,n}))\} I\{x_{i,n} \leq s^\ell\}] \right) \\ & \leq \hat{F}_{X,n}(s^u) - \hat{F}_{X,n}(s^\ell) \\ & \quad + \frac{1}{n} \sum_{i=1}^n \left| F_0(V^u(y^u + a_n + h^u(x_{i,n})) - h_0(x_{i,n})) - F_0(V^\ell(y^\ell + a_n + h^\ell(x_{i,n})) - h_0(x_{i,n})) \right| \\ & \leq 2\epsilon + o(1) \\ & \quad + \frac{1}{n} \sum_{i=1}^n \left| F_0(V^u(y^u + a_n + h^u(x_{i,n})) - h_0(x_{i,n})) - F_0(V^\ell(y^u + a_n + h^\ell(x_{i,n})) - h_0(x_{i,n})) \right| \end{aligned}$$

by (C.3) and the definitions of  $[s^\ell, s^u]$  and  $[y^\ell, y^u]$ . Further, we can bound the last sum by

$$\frac{1}{n} \sum_{i=1}^n \left| F_0(V^u(y^u + a_n + h^u(x_{i,n})) - h_0(x_{i,n})) \right|$$

$$\begin{aligned}
& \left| -F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y^u + a_n + h^u(x_{i,n}))) - h_0(x_{i,n})) \right| \\
& + \frac{1}{n} \sum_{i=1}^n \left| F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y^u + a_n + h^\ell(x_{i,n}))) - h_0(x_{i,n})) \right. \\
& \quad \left. - F_0(V^\ell(y^u + a_n + h^\ell(x_{i,n})) - h_0(x_{i,n})) \right| \\
& + \frac{1}{n} \sum_{i=1}^n \left| F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y^u + a_n + h^u(x_{i,n}))) - h_0(x_{i,n})) \right. \\
& \quad \left. - F_0(\Lambda_0(\Lambda_\vartheta^{-1}(y^u + a_n + h^\ell(x_{i,n}))) - h_0(x_{i,n})) \right| \\
& \leq 3\epsilon
\end{aligned}$$

using the construction of brackets above (note that  $\|V^u - \Lambda_0 \circ \Lambda_\vartheta^{-1}\|_\infty \leq \delta$ ,  $\|\Lambda_0 \circ \Lambda_\vartheta^{-1} - V^\ell\|_\infty \leq \delta$ ,  $\|h^u - h^\ell\|_\infty \leq \gamma$  and recall assumptions **(B5)** and **(B4)**).

Thus  $\sup_{\vartheta \in \Theta} \sup_{\substack{s \in [0,1] \\ y \in C}} |H_{n,\vartheta}^{(1)}(y, s)|$  can be bounded by  $O(\epsilon) + o(1)$  plus a finite maximum over the absolute value of terms

$$\frac{1}{n} \sum_{i=1}^n \left( I\{\Lambda_0(Y_{i,n}) \leq V^u(y^u + a_n + h^u(x_{i,n}))\} - \mathbb{E}[I\{\Lambda_0(Y_{i,n}) \leq V^u(y^u + a_n + h^u(x_{i,n}))\}] \right)$$

and

$$\frac{1}{n} \sum_{i=1}^n \left( I\{\Lambda_0(Y_{i,n}) \leq V^\ell(y^\ell + a_n + h^\ell(x_{i,n}))\} - \mathbb{E}[I\{\Lambda_0(Y_{i,n}) \leq V^\ell(y^\ell + a_n + h^\ell(x_{i,n}))\}] \right).$$

However, those converge to zero in probability by a simple application of Chebychev's inequality.

This completes the proof of  $\sup_{\vartheta \in \Theta} \|H_{n,\vartheta}^{(1)}\| = o_P(1)$  and thus of the lemma.  $\square$

**Lemma C.3** *Under the assumptions of Theorem 4.1 (ii),*

$$\sup_{\vartheta \in \Theta} \|G(\vartheta, h_\vartheta) - G(\vartheta, \hat{h}_\vartheta)\| = o_P(1).$$

**Proof.** The proof is analogous to the proof of Lemma A.3.  $\square$

**Lemma C.4** *Under the assumptions of Theorem 4.1 (ii),*

$$\sup_{\vartheta \in \Theta} \|\tilde{G}_n(\vartheta, \hat{h}_\vartheta) - G(\vartheta, \hat{h}_\vartheta)\| = o_P(1).$$

**Proof.** According to assumption **(N1)** it suffices to show

$$\sup_{\vartheta \in \Theta} \sup_{\substack{s \in [0,1] \\ y \in C}} |\tilde{G}_n(\vartheta, \hat{h}_\vartheta)(y, s) - G(\vartheta, \hat{h}_\vartheta)(y, s)| = o_P(1).$$

Recalling the definitions of  $\tilde{G}_n$  in (C.1) and  $G$  in (3.5) we only consider the first sum and first integral, respectively. It holds by the mean value theorem for integration

$$\begin{aligned}
& \left| \frac{1}{n} \sum_{i=1}^n F_0 \left( \Lambda_0(\Lambda_\vartheta^{-1}(y + \hat{h}_\vartheta(x_{i,n}))) - h_0(x_{i,n}) \right) I\{x_{i,n} \leq s\} \right. \\
& \quad \left. - \int F_0 \left( \Lambda_0(\Lambda_\vartheta^{-1}(y + \hat{h}_\vartheta(x))) - h_0(x) \right) I\{x \leq s\} f_X(x) dx \right| \\
&= \left| \sum_{i=1}^n \left( \frac{1}{n} F_0 \left( \Lambda_0(\Lambda_\vartheta^{-1}(y + \hat{h}_\vartheta(x_{i,n}))) - h_0(x_{i,n}) \right) \right. \right. \\
& \quad \left. \left. - \int_{x_{i-1,n}}^{x_{i,n}} F_0 \left( \Lambda_0(\Lambda_\vartheta^{-1}(y + \hat{h}_\vartheta(x))) - h_0(x) \right) f_X(x) dx \right) I\{x_{i,n} \leq s\} \right. \\
& \quad \left. - \int_{\max\{x_{i,n} | x_{i,n} \leq s\}}^s F_0 \left( \Lambda_0(\Lambda_\vartheta^{-1}(y + \hat{h}_\vartheta(x))) - h_0(x) \right) f_X(x) dx \right| \\
&\leq \sum_{i=1}^n \left| \frac{1}{n} F_0 \left( \Lambda_0(\Lambda_\vartheta^{-1}(y + \hat{h}_\vartheta(x_{i,n}))) - h_0(x_{i,n}) \right) \right. \\
& \quad \left. - F_0 \left( \Lambda_0(\Lambda_\vartheta^{-1}(y + \hat{h}_\vartheta(\xi_{i,n}))) - h_0(\xi_{i,n}) \right) f_X(\xi_{i,n})(x_{i,n} - x_{i-1,n}) \right| \\
& \quad + O(\bar{\Delta}_n)
\end{aligned}$$

for some  $\xi_{i,n} \in [x_{i-1,n}, x_{i,n}]$ . Now the assertion follows from assumption (**A2''**), uniform continuity of  $F_0$  and of  $\Lambda_0 \circ \Lambda_\vartheta^{-1}$  (uniformly in  $\vartheta$ ) and from

$$|\hat{h}_\vartheta(x_{i,n}) - \hat{h}_\vartheta(\xi_{i,n})| \leq \|\hat{h}_\vartheta - h_\vartheta\|_\infty + |\Lambda_\vartheta(\Lambda_0^{-1}(h_0(x_{i,n}))) - \Lambda_\vartheta(\Lambda_0^{-1}(h_0(\xi_{i,n})))|$$

in connection with Lemma 4.2 and assumptions (**A3**), (**B3**).  $\square$

## D Identifiability of the model in the fixed design case

To prove identifiability in the case of deterministic covariates as in Remark 2.4 one starts similarly to the proof in section B of the appendix (main paper) with the cdf of  $\varepsilon_{i,n}(\vartheta_1) = \Lambda_{\vartheta_1}(Y_{i,n}) - h_{\vartheta_1}(x_{i,n})$  in  $y$  to obtain that  $H^{-1}(y + H(h_{\vartheta_0}(x_{i,n}))) - h_{\vartheta_0}(x_{i,n})$  does not depend on  $x_{i,n}$  for  $y \in (-\infty, 0]$ . Due to continuity of the functions and  $\bar{\Delta}_n \rightarrow 0$  one obtains that  $H^{-1}(y + H(h_{\vartheta_0}(x))) - h_{\vartheta_0}(x)$  does not depend on  $x \in [0, 1]$  for  $y \in (-\infty, 0]$ . The remainder of the proof is as in section B.

## References

- Drees, H., Neumeier, N. and Selk, L. (2018). Estimation and hypotheses tests in boundary regression models. *Bernoulli*, to appear.

## E Figures and Tables

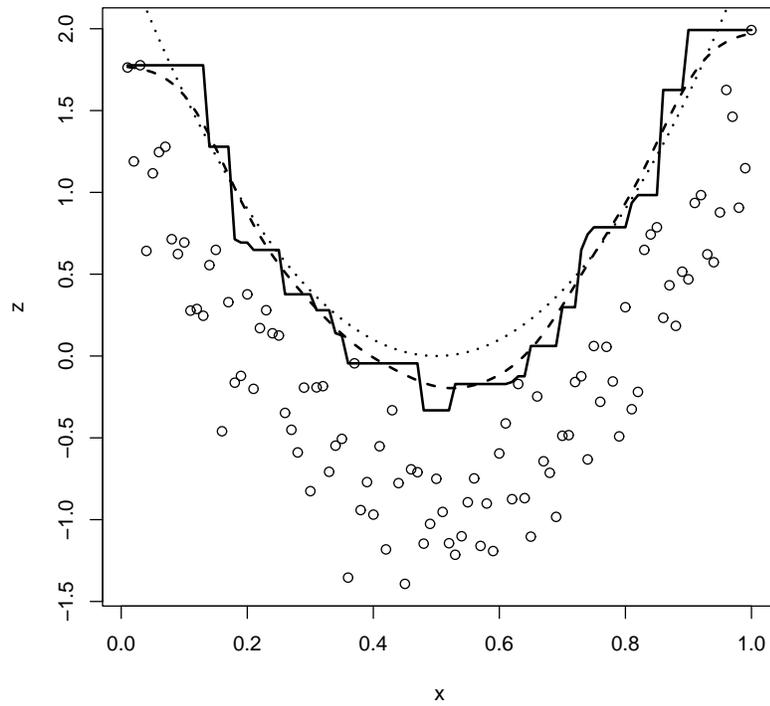


Figure 5: Data corresponding to the model in Figure 1. The true curve is dotted, while the local constant estimator is given by the solid line and the smoothed estimator (with bandwidths  $b_n = n^{-1/3}$  and  $a_n = b_n/2$ ) by the dashed line.

$n = 50$	TKS	TCM	TKSCM	TCMKS
$\vartheta_0 = 0$	0.192 0.197 (0.102)	<b>0.001</b> 0.009 ( <u>0.037</u> )	-0.038 -0.139 (0.118)	0.225 0.208 (0.118)
$\vartheta_0 = 0.5$	0.778 0.691 (0.191)	<b>0.378</b> 0.402 ( <u>0.092</u> )	0.239 0.302 (0.410)	0.858 0.798 (0.274)
$\vartheta_0 = 1$	1.290 1.340 (0.233)	<b>0.728</b> 0.741 ( <u>0.232</u> )	0.388 0.264 (1.000)	1.370 1.350 (0.308)
$\vartheta_0 = 1.5$	<b>1.750</b> 1.790 ( <u>0.195</u> )	1.160 1.290 (0.368)	0.507 0.292 (1.810)	1.790 1.790 (0.222)
$\vartheta_0 = 2$	1.940 2.060 (0.201)	1.590 1.750 (0.478)	0.585 0.424 (2.880)	<b>1.970</b> 2.060 ( <u>0.141</u> )

$n = 100$	TKS	TCM	TKSCM	TCMKS
$\vartheta_0 = 0$	<b>0.017</b> 0.018 (0.037)	0.080 0.079 ( <u>0.014</u> )	0.061 0.073 (0.022)	-0.020 -0.004 (0.020)
$\vartheta_0 = 0.5$	<b>0.496</b> 0.517 ( <u>0.028</u> )	0.338 0.346 (0.042)	0.516 0.578 (0.080)	0.521 0.548 (0.032)
$\vartheta_0 = 1$	0.973 0.979 ( <u>0.044</u> )	0.745 0.745 (0.092)	0.906 1.050 (0.225)	<b>1.030</b> 1.020 (0.054)
$\vartheta_0 = 1.5$	1.480 1.460 ( <u>0.059</u> )	1.210 1.230 (0.123)	1.310 1.510 (0.412)	<b>1.510</b> 1.490 (0.060)
$\vartheta_0 = 2$	<b>1.960</b> 2.000 (0.059)	1.690 1.740 (0.144)	1.550 1.860 (0.822)	1.920 1.940 ( <u>0.058</u> )

Table 6: Mean, median and MISE for Model (5.2) for  $n = 50$  and  $n = 100$  with  $a_n = b_n/2$ .

$n = 50$	TKS	TCM	TKSCM	TCMKS
$\vartheta_0 = 0$	0.156 0.167 (0.076)	-0.050 -0.062 ( <u>0.031</u> )	<b>0.022</b> 0.103 (0.095)	0.191 0.198 (0.086)
$\vartheta_0 = 0.5$	0.713 0.646 (0.130)	<b>0.324</b> 0.336 ( <u>0.088</u> )	0.268 0.407 (0.348)	0.781 0.695 (0.197)
$\vartheta_0 = 1$	1.260 1.310 ( <u>0.191</u> )	0.655 0.646 (0.242)	0.447 0.511 (0.919)	<b>1.330</b> 1.350 (0.258)
$\vartheta_0 = 1.5$	<b>1.720</b> 1.780 (0.188)	1.100 1.180 (0.365)	0.619 0.559 (1.660)	<b>1.720</b> 1.780 ( <u>0.177</u> )
$\vartheta_0 = 2$	<b>1.970</b> 2.060 (0.141)	1.550 1.660 (0.442)	0.726 0.619 (2.630)	1.960 2.060 ( <u>0.111</u> )

$n = 100$	TKS	TCM	TKSCM	TCMKS
$\vartheta_0 = 0$	<b>0.001</b> 0.050 (0.044)	0.129 0.128 (0.023)	0.028 0.037 ( <u>0.016</u> )	-0.014 -0.042 ( <u>0.015</u> )
$\vartheta_0 = 0.5$	0.467 0.474 ( <u>0.033</u> )	0.282 0.287 (0.063)	<b>0.497</b> 0.533 (0.057)	0.481 0.486 ( <u>0.034</u> )
$\vartheta_0 = 1$	0.934 0.942 ( <u>0.043</u> )	0.674 0.649 (0.130)	0.878 0.999 (0.190)	<b>0.965</b> 0.960 (0.049)
$\vartheta_0 = 1.5$	1.420 1.390 (0.056)	1.120 1.130 (0.185)	1.320 1.500 (0.336)	<b>1.440</b> 1.400 ( <u>0.053</u> )
$\vartheta_0 = 2$	<b>1.910</b> 1.920 ( <u>0.071</u> )	1.590 1.610 (0.228)	1.560 1.850 (0.790)	1.850 1.790 (0.079)

Table 7: Mean, median and MISE for Model (5.2) for  $n = 50$  and  $n = 100$  with  $a_n = b_n/20$ .

$n = 50$	TKS	TCM	TKSCM	TCMKS
$\vartheta_0 = 0$	0.141 0.010 (0.309)	<b>0.020</b> 0.012 ( <u>0.008</u> )	-0.056 -0.053 (0.016)	0.137 0.028 (0.260)
$\vartheta_0 = 0.5$	0.519 0.549 (0.039)	<b>0.518</b> 0.506 ( <u>0.023</u> )	0.521 0.532 (0.040)	0.546 0.574 (0.038)
$\vartheta_0 = 1$	1.010 1.010 (0.085)	<b>1.000</b> 0.996 ( <u>0.040</u> )	0.996 0.998 (0.071)	1.040 1.030 (0.077)
$\vartheta_0 = 1.5$	1.530 1.530 (0.125)	<b>1.500</b> 1.490 ( <u>0.066</u> )	<b>1.500</b> 1.510 (0.113)	1.550 1.570 (0.110)
$\vartheta_0 = 2$	1.960 2.060 (0.118)	<b>2.010</b> 2.040 ( <u>0.069</u> )	1.950 2.000 (0.156)	1.970 2.050 (0.093)

$n = 100$	TKS	TCM	TKSCM	TCMKS
$\vartheta_0 = 0$	0.019 0.009 (0.022)	<b>0.006</b> 0.000 ( <u>0.004</u> )	0.043 0.038 (0.007)	-0.014 -0.007 (0.013)
$\vartheta_0 = 0.5$	0.522 0.524 (0.023)	<b>0.505</b> 0.498 ( <u>0.013</u> )	0.562 0.555 (0.020)	0.528 0.524 (0.022)
$\vartheta_0 = 1$	1.030 1.030 (0.042)	<b>1.010</b> 1.000 ( <u>0.021</u> )	1.080 1.080 (0.038)	1.030 1.020 (0.042)
$\vartheta_0 = 1.5$	1.550 1.550 (0.061)	<b>1.510</b> 1.510 ( <u>0.030</u> )	1.600 1.590 (0.055)	1.550 1.550 (0.061)
$\vartheta_0 = 2$	2.040 2.060 (0.066)	<b>2.000</b> 2.000 ( <u>0.037</u> )	2.070 2.070 (0.061)	2.020 2.050 (0.058)

Table 8: Mean, median and MISE for Model (5.3) for  $n = 50$  and  $n = 100$  with  $a_n = b_n/2$ .

$n = 50$	TKS	TCM	TKSCM	TCMKS
$\vartheta_0 = 0$	0.097 0.031 (0.223)	<b>0.005</b> 0.016 ( <u>0.008</u> )	-0.037 -0.030 (0.013)	-0.087 -0.001 (0.174)
$\vartheta_0 = 0.5$	0.487 0.506 (0.039)	0.479 0.462 ( <u>0.021</u> )	<b>0.506</b> 0.508 (0.036)	0.514 0.522 (0.035)
$\vartheta_0 = 1$	0.976 0.978 (0.092)	0.965 0.962 ( <u>0.044</u> )	<b>0.984</b> 0.997 (0.074)	1.020 1.020 (0.078)
$\vartheta_0 = 1.5$	<b>1.499</b> 1.469 (0.120)	1.440 1.430 ( <u>0.063</u> )	1.450 1.470 (0.119)	1.530 1.500 (0.105)
$\vartheta_0 = 2$	1.920 1.990 (0.105)	<b>1.960</b> 1.960 ( <u>0.069</u> )	1.940 1.940 (0.127)	1.930 1.970 (0.086)

$n = 100$	TKS	TCM	TKSCM	TCMKS
$\vartheta_0 = 0$	0.017 0.004 (0.016)	<b>0.004</b> 0.000 ( <u>0.004</u> )	0.042 0.039 (0.007)	-0.010 -0.007 (0.007)
$\vartheta_0 = 0.5$	0.530 0.537 (0.021)	<b>0.507</b> 0.497 ( <u>0.011</u> )	0.563 0.554 (0.019)	0.534 0.539 (0.021)
$\vartheta_0 = 1$	1.020 1.020 (0.042)	<b>1.000</b> 1.000 ( <u>0.020</u> )	1.080 1.070 (0.035)	1.020 1.010 (0.039)
$\vartheta_0 = 1.5$	1.550 1.560 (0.064)	<b>1.510</b> 1.510 ( <u>0.031</u> )	1.600 1.600 (0.054)	1.560 1.550 (0.064)
$\vartheta_0 = 2$	2.050 2.060 (0.069)	<b>2.020</b> 2.040 ( <u>0.041</u> )	2.090 2.100 (0.064)	2.030 2.060 (0.059)

Table 9: Mean, median and MISE for Model (5.3) for  $n = 50$  and  $n = 100$  with  $a_n = b_n/20$ .

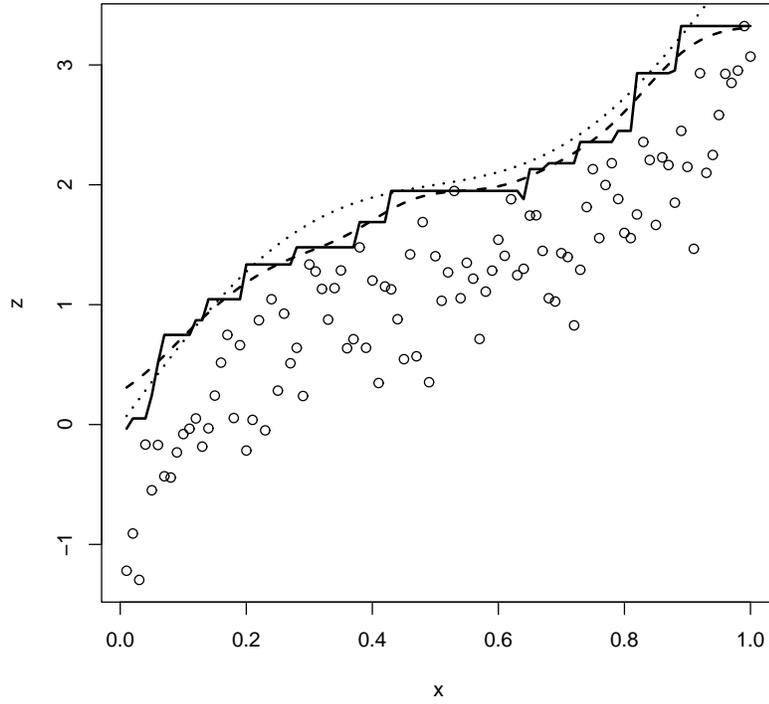


Figure 6: The setting is similar to Figure 5 with model from Figure 2 in the main paper.

Method	Pearson	Kendall	Spearman
Original data	-0.634	-0.456	-0.612
True parameter $\vartheta_0$	0.001	0.001	0.001
TKS	0.001	0.001	0.001
TCM	0.009	0.006	0.008
TKSCM	0.005	0.002	0.004
TCMKS	0.002	0.001	0.001

Table 10: *Pearson's, Kendall's and Spearman's correlation coefficients (the average over 1000 iterations) between the covariates and the errors for the model (5.3) when  $n = 100$ . The first line corresponds to the correlations for the original data while the second line is for the true transformation parameter ( $\vartheta_0 = 0.5$ ). The last four lines correspond to the correlations for each estimator.*