

van Heel, Mareike; Dikta, Gerhard; Braekers, Roel

**Article — Published Version**

## Bootstrap based goodness-of-fit tests for binary multivariate regression models

Journal of the Korean Statistical Society

**Provided in Cooperation with:**

Springer Nature

*Suggested Citation:* van Heel, Mareike; Dikta, Gerhard; Braekers, Roel (2021) : Bootstrap based goodness-of-fit tests for binary multivariate regression models, Journal of the Korean Statistical Society, ISSN 2005-2863, Springer, Singapore, Vol. 51, Iss. 1, pp. 308-335, <https://doi.org/10.1007/s42952-021-00142-4>

This Version is available at:

<https://hdl.handle.net/10419/287618>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>



# Bootstrap based goodness-of-fit tests for binary multivariate regression models

Mareike van Heel<sup>1,2</sup> · Gerhard Dikta<sup>1</sup> · Roel Braekers<sup>3</sup>

Received: 7 September 2020 / Accepted: 27 July 2021 / Published online: 4 August 2021  
© The Author(s) 2021

## Abstract

We consider a binary multivariate regression model where the conditional expectation of a binary variable given a higher-dimensional input variable belongs to a parametric family. Based on this, we introduce a model-based bootstrap (MBB) for higher-dimensional input variables. This test can be used to check whether a sequence of independent and identically distributed observations belongs to such a parametric family. The approach is based on the empirical residual process introduced by Stute (Ann Statist 25:613–641, 1997). In contrast to Stute and Zhu's approach (2002) Stute & Zhu (Scandinavian J Statist 29:535–545, 2002), a transformation is not required. Thus, any problems associated with non-parametric regression estimation are avoided. As a result, the MBB method is much easier for users to implement. To illustrate the power of the MBB based tests, a small simulation study is performed. Compared to the approach of Stute & Zhu (Scandinavian J Statist 29:535–545, 2002), the simulations indicate a slightly improved power of the MBB based method. Finally, both methods are applied to a real data set.

**Keywords** Binary regression model · Bootstrap based test · Goodness-of-fit test · Marked empirical process

**Mathematics subject classification** 60F17 · 62E20 · 62F40 · 62J12

---

✉ Mareike van Heel  
van-heel@fh-aachen.de

Gerhard Dikta  
dikta@fh-aachen.de

Roel Braekers  
roel.braekers@uhasselt.be

<sup>1</sup> Fachhochschule Aachen, D-52428 Jülich, Germany

<sup>2</sup> Universiteit Hasselt, B-3500 Hasselt, Belgium

<sup>3</sup> Data Science Institute, I-Biostat, Universiteit Hasselt, B-3500 Hasselt, Belgium

## 1 Introduction

Binary multivariate regression models are for example used to analyze longitudinal data. Those appear in clinical studies and are used to evaluate the effect of interventions over time. For different individuals, information is collected at several assessment times. To deal with incomplete data in a longitudinal setup, inverse probability weighted generalized estimating equations (WGEE) (Robins et al., 1994) are used. The resulting WGEE provides consistent estimators only if the underlying (binary) process of missing data is properly modeled. Of course, this should be secured in advance.

This paper addresses, in a more general context than just described, the question of how to test the model assumptions of a binary generalized linear regression model.

Mathematically, we describe the data with a sequence of independent and identically distributed (iid) random variables

$$(\delta_1, X_1), \dots, (\delta_n, X_n),$$

where  $\delta$  is a binary or 0–1 response variable and  $X \in \mathbb{R}^d$  a  $d$ -dimensional input with continuous distribution function (df)  $H$ . For the binary regression model,

$$\hat{m} : \mathbb{R}^d \ni x \rightarrow \hat{m}(x) = \mathbb{E}(\delta|X = x) \equiv \mathbb{P}(\delta = 1|X = x) \in [0, 1]$$

denotes the conditional expectation of  $\delta$  given  $X = x$ . Under the generalized linear model (GLM), one assumes that there exists a link function  $g$ , that is an invertible function with measurable inverse, such that

$$g(\mathbb{E}(\delta|X = x)) = \beta_0^\top x,$$

for  $H$  almost all  $x \in \mathbb{R}^d$  and an appropriate  $\beta_0 \in \mathbb{R}^d$ . The function  $g$  is assumed to be known. Based on this, we set

$$m : \mathbb{R} \ni t \rightarrow m(t) = g^{-1}(t) \in [0, 1].$$

Assuming that the data  $(\delta, X)$  comes from a GLM with link function  $g$  now means that  $\hat{m} \in M := \{m(\beta^\top \cdot) | \beta \in \mathbb{R}^d\}$ .

If one assumes a GLM to analyze a sample  $(\delta_1, X_1), \dots, (\delta_n, X_n)$  of iid data, one has to guarantee that the linear part and the assumed link-function are correct or, at least, that the data shows no obvious departure from the model. Thus, we need a goodness-of-fit test to validate the model, i. e., we need a universal test to check the null hypothesis

$$H_0 : \hat{m} \in M \text{ versus } H_1 : \hat{m} \notin M.$$

A general approach for model checking in a regression setup was introduced by Stute, (1997). Stute & Zhu, (2002) specialized this approach to GLM, where the response variable is not necessarily binary. In the binary setup of GLM, the underlying probabilistic background is a functional limit result of the marked empirical process with estimated parameters:

$$R_n^1(t) = n^{-1/2} \sum_{i=1}^n (\delta_i - m(\beta_n^\top X_i)) I(\beta_n^\top X_i \leq t), \quad t \in \mathbb{R},$$

where  $\beta_n$  is a proper estimator of  $\beta_0$  and  $I$  denotes the indicator function, see Stute (1997). With  $R_n^1(-\infty) = 0$  and  $R_n^1(\infty) = n^{-1/2} \sum_{i=1}^n (\delta_i - m(\beta_n^\top X_i))$ , this process is a random element in the Skorokhod space  $D([-\infty, \infty])$ . Under appropriate conditions,  $R_n^1$  converges in distribution against a centered Gaussian process  $R_\infty^1$ , which, however, has a rather complicated, model-dependent covariance structure, cf. Theorem 1 in Stute & Zhu, (2002). To make this result usable for applications in statistics, Stute and Zhu introduced a model-based transformation. Applying this transformation, respectively its estimated version, to  $R_n^1$ , this composition converges in distribution against a time-transformed Brownian motion, cf. Theorem 2 in Stute & Zhu, (2002). This framework is then used to get asymptotically distribution-free statistics.

The approach works excellently, but has two weak points. For the transformation, one needs an estimate of the conditional expectation of  $X$  given  $\beta_0^\top X = v$ ,  $\mathbb{E}(X | \beta_0^\top X = v)$ , for all  $v \in \mathbb{R}$ . Under general conditions, one must estimate this quantity using a non-parametric procedure. However, such a method always requires a smoothing parameter, but its selection is not unproblematic. However, since the model as a whole is parametric, the question inevitably arises whether this non-parametric method is absolutely necessary. Moreover, a user who wants to check a chosen GLM with this method must implement the model-dependent transformation in each case. This is of course feasible, but goes along with a considerable effort, because the transformation is quite complex especially for non-statisticians. Of course, parts of this procedure could be automated and implemented as software, but then it will hardly be applicable without the appropriate knowledge about the transformation. It would be nice if all this could be avoided.

To estimate  $\beta_0$ , we use the maximum likelihood estimator (MLE) given by

$$\beta_n = \arg \max_{\beta \in \mathbb{R}^d} l_n(\beta),$$

where

$$l_n(\beta) = \frac{1}{n} \sum_{i=1}^n (\delta_i \ln(m(\beta^\top X_i)) + (1 - \delta_i) \ln(1 - m(\beta^\top X_i)))$$

is the normalized log-likelihood function.

For the bootstrap data, we propose the following model-based (MB) resampling scheme similar to the resampling scheme in Dikta et al., (2006). MBB guarantees that the bootstrap data are always generated according to the null hypothesis.

**Definition 1** Let  $(\delta_1, X_1), \dots, (\delta_n, X_n)$  be iid observations, where the  $\delta_i$  are binary and the  $X_i$  have a continuous distribution function  $H$ . Let  $\beta_n$  be the corresponding MLE. The model-based resampling scheme is then defined as follows:

1. Set  $X_i^* = X_i$  for  $1 \leq i \leq n$ .

2. Generate a sample  $\delta_1^*, \dots, \delta_n^*$  of independent Bernoulli random variables where  $\delta_i^*$  has the probability of success given by  $m(\beta_n^\top X_i)$ , for  $1 \leq i \leq n$ , where  $m(\beta^\top x) = \mathbb{P}_\beta(\delta = 1 | X = x)$ .

Under this resampling scheme, only the  $\delta$ 's are resampled, the corresponding  $X$ 's are taken from the original sample.

To define  $R_n^{1*}$ , the bootstrap analog of  $R_n^1$ , we assume a bootstrap sample

$$(\delta_1^*, X_1^*), \dots, (\delta_n^*, X_n^*)$$

and set

$$R_n^{1*}(t) = n^{-1/2} \sum_{i=1}^n (\delta_i^* - m(\beta_n^{*\top} X_i^*)) I(\beta_n^\top X_i^* \leq t), \quad t \in \mathbb{R},$$

where  $\beta_n^*$  is the MLE corresponding to the log-likelihood function based on the bootstrap sample. Usually the  $\beta_n$  in the indicator is also replaced by  $\beta_n^*$ . We don't replace it here, since both processes can be shown to be asymptotically equivalent. Furthermore, simulations that use  $\beta_n$  instead of  $\beta_n^*$  run faster, since  $\beta_n$  is the same for each bootstrap sample.

We will prove that the cumulative residual process  $R_n^{1*}(t)$  corresponding to the MB bootstrap data behaves asymptotically as  $R_n^1$  if the original data satisfy the null hypothesis. Thus, the distribution of any statistic that depends continuously on  $R_n^1$  can be approximated by the corresponding distribution based on  $R_n^{1*}$ . This provides the basic asymptotic backup of our method. But in addition to this, an approximation based on  $R_n^{1*}$  also has the advantage that it accurately reflects the fixed sample sizes. Even if the original data come from the alternative, the bootstrap data are always generated under the null hypothesis. Thus, a statistic based on  $R_n^{1*}$  fits the null hypothesis. This is crucial because p-values are based on the distribution under the null hypothesis. Overall, this should lead to a more accurate approximation of the p-values compared to the pure asymptotic one under finite sample size, and, hence, to an improvement of the power. Indeed, we can observe some improvements in the simulation study.

As in Stute & Zhu, (2002), we consider a Kolmogorov-Smirnov (KS) and Cramér-von Mises (CvM) test statistics  $D_n$  and  $W_n$  based on  $R_n^1$  as

$$D_n = \sup_{t \in \mathbb{R}} |R_n^1(t)|$$

and

$$W_n = \int (R_n^1(t))^2 H_n(dt).$$

Here  $H_n$  is the empirical distribution function (edf) of the  $\beta_n^\top X$  sample. Since, under  $H_0$ ,  $R_n^1 \rightarrow R_\infty^1$  in distribution, as  $n \rightarrow \infty$ , the continuous mapping theorem implies that

$$D_n \longrightarrow D_\infty \equiv \sup_{t \in \mathbb{R}} |R_\infty^1(t)|,$$

and  $W_n \rightarrow W_\infty$  in distribution, as  $n \rightarrow \infty$ .

If under  $H_0$  the process  $R_n^{1*}$  tends in distribution to the same limiting process  $R_\infty^1$  as  $R_n^1$ , the p-values corresponding to the KS and CvM test can now be approximated by the typical Monte-Carlo approach (used in bootstrap applications) based on the distribution of

$$D_n^* = \sup_{t \in \mathbb{R}} |R_n^{1*}(t)|$$

and

$$W_n^* = \int (R_n^{1*}(t))^2 H_n^*(dt),$$

where  $H_n^*$  denotes the edf based on the  $\beta_n^T X^*$  sample.

This article is organized as follows: In Section 2 we state the main results, which guarantee that the MBB can be used to test our null hypothesis. In Section 3 the approach is applied in a simulation study and a real data application. Here, our approach is also compared to the approach by Stute & Zhu, (2002). The results of Section 3 are discussed in Section 4. The proofs of our main results are provided in Section 5. Additionally, in the Appendix some results used in Section 5 are presented.

## 2 Main results

In this chapter, our main result is given in Theorem 2.

To prove Theorem 2, we first show that in the space  $D[-\infty, \infty]$ , the process  $R_n^*(u) = n^{-1/2} \sum_{i=1}^n (\delta_i^* - m(\beta_n^T X_i)) I(\beta_n^T X_i \leq u) \rightarrow R_\infty$  in distribution, where  $R_\infty$  is a centered Gaussian process, see Theorem 1. This process is similar to  $R_n^1(u)$ , but  $\delta$  is replaced with  $\delta^*$ . Theorem 1 is a stepping stone for proving Theorem 2, in which we also replace  $\beta_n$  with  $\beta_n^*$ . To prove both theorems, we show that the fidis of both processes converge and that the processes are tight, see Theorem 13.5 of Billingsley, (1999). Lemma 1 (iii) provides a result which is required to prove the convergence of the fidis of the process  $R_n^*$ . Lemma 1 (i) and Lemma 1 (ii) are required to prove Lemma 1 (iii).

Since we finally replace  $\beta_n$  with  $\beta_n^*$  in Theorem 2, we need to ensure that  $\beta_n^*$  converges to  $\beta_n$ , which is done in Lemma 2. The proof of Theorem 2 uses a decomposition of the process  $R_n^{1*}(u)$  into  $R_n^*(u)$  and a difference term. To simplify the representation, Lemma 3 is used. With the final decomposition we now prove the tightness and the convergence of the fidis of the process  $R_n^{1*}(u)$ .

For Theorem 1 we need the following assumptions:

(A1)  $\beta_n \rightarrow \beta_0$ , as  $n \rightarrow \infty$ , w.p. 1.

(B1) Define

$$H(u, \beta) = \int m(\beta^T X) \bar{m}(\beta^T X) I(\beta^T X \leq u) d\mathbb{P},$$

where  $\bar{m} = 1 - m$ .  $H$  is uniformly continuous in  $u$  at  $\beta$ .

(C1)  $m(\beta^T x)$  is continuous in  $\beta^T x$ .

(D1)  $m(\beta^T x)$  is continuous differentiable in  $\beta^T x$  with  $m'(\beta^T x) = \partial m(\beta^T x) / \partial(\beta^T x)$  and  $m'$  is bounded.

Assumptions (C1), (D1) and (B1) are similar to assumptions (B) and (C) in Stute & Zhu, (2002), but specified to the binary setup. Furthermore, with (A1) we ensure that  $\beta_n \rightarrow \beta_0$ , as  $n \rightarrow \infty$ , w.p. 1.

As mentioned before, the following Lemma is used to prove the convergence of the fidis of the process  $R_n^*(u)$ , which is defined in Theorem 1.

**Lemma 1** (i) *If assumption (D1) is fulfilled,*

$$\sup_{\beta \in \mathbb{R}^d, u \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n I(\beta^T X_i \leq u) m(\beta^T X_i) \bar{m}(\beta^T X_i) - \mathbb{E}(I(\beta^T X \leq u) m(\beta^T X) \bar{m}(\beta^T X)) \right| \rightarrow 0,$$

as  $n \rightarrow \infty$ , w.p. 1.

(ii) *If assumptions (B1) and (C1) are fulfilled,*

$$\sup_{|\beta - \beta_0| \leq \varepsilon, u \in \mathbb{R}} \left| \mathbb{E}(I(\beta^T X \leq u) m(\beta^T X) \bar{m}(\beta^T X)) - \mathbb{E}(I(\beta_0^T X \leq u) m(\beta_0^T X) \bar{m}(\beta_0^T X)) \right| \rightarrow 0,$$

as  $\varepsilon \rightarrow 0$ .

(iii) *If assumptions (A1), (B1), (C1) and (D1) are fulfilled and  $\mathbb{E}(|X|) < \infty$ , then*

$$\sup_{u \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n I(\beta_n^T X_i \leq u) m(\beta_n^T X_i) \bar{m}(\beta_n^T X_i) - \mathbb{E}(I(\beta_0^T X \leq u) m(\beta_0^T X) \bar{m}(\beta_0^T X)) \right| \rightarrow 0,$$

as  $n \rightarrow \infty$ , w.p. 1.

Now, in the process  $R_n^1(u)$ , we replace  $\delta$  with  $\delta^*$ , where  $\delta^*$  is generated by using the MB scheme. As stated in the following Theorem,  $R_n^*(u)$  converges.

**Theorem 1** Assume that  $\mathbb{E}(|X|) < \infty$ , assumptions (A1), (B1), (C1) and (D1) are satisfied, and the MB resampling scheme is used to generate the bootstrap data, then, w.p. 1, under the null hypothesis, the process

$$R_n^*(u) = n^{-1/2} \sum_{i=1}^n (\delta_i^* - m(\beta_n^T X_i)) I(\beta_n^T X_i \leq u) \rightarrow R_\infty$$

in distribution in the space  $D[-\infty, \infty]$ , where  $R_\infty$  is a centered Gaussian process with covariance function

$$K(s, t) = \mathbb{E}(R_\infty(s), R_\infty(t)) = \int m(\beta_0^T X) \bar{m}(\beta_0^T X) I(\beta_0^T X \leq s \wedge t) d\mathbb{P}.$$

After replacing  $\delta$  with  $\delta^*$ , we need to replace  $\beta_n$  with  $\beta_n^*$ .

For this, we define

$$\begin{aligned} l(\beta^T X, \delta^*) &= \frac{\partial}{\partial \beta} (\delta^* \ln(m(\beta^T X)) + (1 - \delta^*) \ln(1 - m(\beta^T X))) \\ &= \delta^* \frac{w(X, \beta)}{m(\beta^T X)} - (1 - \delta^*) \frac{w(X, \beta)}{1 - m(\beta^T X)} \end{aligned}$$

which is the derivative of the summands of the log-likelihood function and  $w(x, \beta) = \partial m(\beta^T x) / \partial \beta = (w_1(x, \beta), \dots, w_d(x, \beta))^T$ .

Check that  $\mathbb{E}_n^*(l(\beta_n^T X, \delta^*)) = 0$ .

For the following Lemmas and Theorem 2 we need some additional assumptions:

- (A2)  $L(\beta_0) = \mathbb{E}(l(\beta_0^T X, \delta) l^T(\beta_0^T X, \delta))$  exists and is positive definite.
- (B2)  $n^{1/2}(\beta_n^* - \beta_n) = n^{-1/2} \sum_{i=1}^n l(\beta_n^T X_i, \delta_i^*) + o_{\mathbb{P}^*}(1)$ , w.p. 1.
- (C2)  $L_n^*(\beta_n) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_n^*(l(\beta_n^T X_i, \delta_i^*) l^T(\beta_n^T X_i, \delta_i^*)) \rightarrow L(\beta_0)$ , w.p. 1.
- (D2) For every  $x \in \mathbb{R}^d$ ,  $w(x, \beta) = \partial m(\beta^T x) / \partial \beta = (w_1(x, \beta), \dots, w_d(x, \beta))^T$  exists and is continuous with respect to  $\beta$  for every  $\beta$  in a neighborhood of  $\beta_0$  (not depending on  $x$ ).
- (E2) There exists a square-integrable function  $M(x)$  such that for every  $x$ 

$$\max \left( \frac{w_i(x, \beta)}{m(\beta^T x)}, \frac{w_i(x, \beta)}{1 - m(\beta^T x)} \right) \leq M(x)$$
for every  $\beta$  in a neighborhood of  $\beta_0$  and  $1 \leq i \leq d$ .
- (F2) The function

$$W : \mathbb{R} \times V_\beta \ni (x, \beta) \rightarrow W(x, \beta) = \mathbb{E}(w(X, \beta_0) I(\beta^T X \leq x)) \in \mathbb{R}^d$$

is uniformly continuous in  $u$  at  $\beta_0$ , where  $V_\beta = \{\beta : \beta \in V\}$  and  $V$  is given under (D2).

Assumptions (D2) and (E2) are again similar to assumption (B) in Stute & Zhu, (2002), but specified to the binary setup. Furthermore, assumptions (A2) and (B2) are similar to assumption (A).

Lemma 2 is necessary to ensure that  $\beta_n^*$  converges to  $\beta_n$ .



**Lemma 2** Assume that assumptions (A1), (A2), (B2), (C2) and (E2) hold. Then, w.p. 1,

$$n^{1/2}(\beta_n^* - \beta_n) \rightarrow Z, \text{ as } n \rightarrow \infty,$$

where  $Z$  is a multivariate normal distribution with zero mean and covariance matrix  $L(\beta_0)$ .

In addition, we need some results for  $w(x, \beta)$  and  $W(x, \beta)$ .

**Lemma 3** Let  $\hat{\beta}_n^* : \mathbb{R}^d \rightarrow V$  be a measurable function such that  $\hat{\beta}_n^*(x)$  lies in the line segment that connects  $\beta_n^*$  and  $\beta_n$  for each  $x \in \mathbb{R}^d$  and assume (A1), (A2), (D2), (E2) and (F2) hold. Then, w.p. 1, for  $1 \leq j \leq d$ ,

- (i)  $\sup_{u \in \mathbb{R}} \left| n^{-1} \sum_{i=1}^n w_j(X_i, \beta_0) I(\beta_n^{*T} X_i \leq u) - W_j(u, \beta_0) \right| \rightarrow 0, \text{ as } n \rightarrow \infty,$
- (ii)  $\sup_{u \in \mathbb{R}} \left| n^{-1} \sum_{i=1}^n (w_j(X_i, \hat{\beta}_n^*(X_i)) - w_j(X_i, \beta_0)) I(\beta_n^{*T} X_i \leq u) \right| = o_{\mathbb{P}_n^*}(1).$

Finally, the process  $R_n^{1*}(u)$  converges in distribution.

**Theorem 2** Assume that  $\mathbb{E}(|X|) < \infty$ , assumptions (A1), (B1), (C1), (D1), (A2), (B2), (C2), (D2), (E2) and (F2) are satisfied, and the MB resampling scheme is used to generate the bootstrap data, then, w.p. 1, under the null hypothesis, the process

$$R_n^{1*}(u) = n^{-1/2} \sum_{i=1}^n (\delta_i^* - m(\beta_n^{*T} X_i)) I(\beta_n^{*T} X_i \leq u) \rightarrow R_\infty^1$$

in distribution in the space  $D[-\infty, \infty]$ , where  $R_\infty^1$  is a centered Gaussian process with covariance function

$$\hat{K}(s, t) = K(s, t) + W^T(s, \beta_0)L(\beta_0)W(t, \beta_0) - 2W^T(s, \beta_0)W(t, \beta_0).$$

### 3 Simulations and real data application

#### 3.1 Simulations

To clarify the results, the Bootstrap approach is compared to the approach introduced by Stute & Zhu, (2002). For the application of their method, we make use of an additional assumption to avoid the non-parametric estimation of  $\mathbb{E}(X|\beta_0^T X = \nu)$ . As stated in Stute & Zhu, (2002), page 541, we assume that  $X$  belongs to a family of elliptically contoured distributions. Note that we do not need this assumption for our bootstrap approach. To calculate the p-values for the approach by Stute and Zhu we

use the Karhunen-Loève expansion for a Brownian motion [(Bass, 2011), formula (6.2)] to approximate the distribution of the integrated squared Brownian motion over the unit interval.

In all simulations, the empirical powers and ecdfs of the p-values based on the CvM statistic are calculated from 1000 replications. The sample sizes are set to  $n = 50$  and  $n = 100$ . For the Bootstrap approach, each p-value is based on 200 bootstrap samples. The ecdfs of the 1000 p values per simulation and approach are displayed in a graph together with the uniform distribution function (red: Bootstrap approach, blue: approach by Stute and Zhu, gray: uniform distribution function). In addition, the percentages of rejecting the null hypothesis (at levels  $\alpha = 0.05$  and  $\alpha = 0.01$ ) are given explicitly.

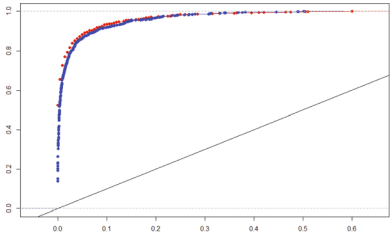
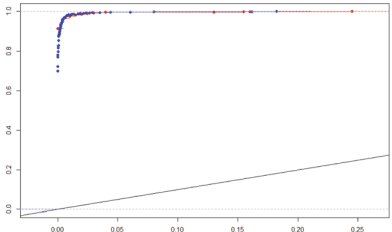
In the first simulation, we generate uncorrelated  $X_i$  from a 3-dimensional normal distribution with mean values 0 and variance 1. Based on a chosen  $\beta$  ( $\beta = (1, 1, 2)^T$ ) we calculate the probability  $P(\delta = 1|X = x)$ , assuming a logistic regression model. In our test, we assume that the generated data belong to a GLM with a logistic regression function where  $\beta$  is 3-dimensional, which is true. Table 1 shows the results. The two ecdfs of the p values based on the CvM statistic are very similar to the distribution function of a uniform distribution. Thus, the test holds the level.

In the second simulation, the data are generated the same way as in the first simulation, but now the third covariate was squared. We assume that the data belong to a GLM with a logistic regression function where the third component is not squared, which is false. Table 2 shows that both approaches yield similar results. Furthermore, in both cases the power increased with the sample size.

**Table 1**  $H_0$  (logistic regression function,  $\beta$  is 3-dimensional) is true

		BS	Stute/Zhu	Ecdfs (red: BS, blue: Stute/ Zhu)
$n = 50$	$\alpha = 0.05$	5.2%	3.8%	
	$\alpha = 0.01$	0.8%	0.4%	
$n = 100$	$\alpha = 0.05$	5.5%	4.9%	
	$\alpha = 0.01$	1.4%	0.4%	

**Table 2**  $H_0$  (logistic regression function,  $\beta$  is 3-dimensional, third component not squared) is false

		Bootstrap	Stute/Zhu	Ecdfs (red: BS, blue: Stute/ Zhu)
$n = 50$	$\alpha = 0.05$	87.4%	86.3%	
	$\alpha = 0.01$	65.5%	65.9%	
$n = 100$	$\alpha = 0.05$	99.6%	99.6%	
	$\alpha = 0.01$	96.3%	98.3%	

In the third simulation, we generate data using a nonparametric mixing of logistic regression, see Agresti, (2002), 13.2.2. The  $X_i$  are again generated from a 3-dimensional normal distribution with mean values 0 and variance 1 and  $\beta = (1, 1, 2)^T$ . Furthermore, a Bernoulli variable with  $p = 0.2$  is generated. If this variable is 0, we add 1 to  $\beta^T X$ . Again we calculate the probability  $P(\delta = 1|X = x)$  assuming a logistic regression model. In our test, we assume that the generated data belong to a GLM with a logistic regression function where  $\beta$  is 3-dimensional, which is false. Table 3 shows similar results as in the second simulation.

In the last simulation, we generate the data in the same way as in the first simulation again. This time, we assume a probit regression model where  $\beta$  is 3-dimensional, which is false. Table 4 shows, that all ecdfs of the  $p$  values based on the CvM statistic are very similar to the distribution function of a uniform distribution. Thus, both tests do not detect this departure from the null hypothesis.

### 3.2 Real data application

We applied the introduced test to the data set reported on by Härdle and Stoker, (1989). This data set consists of 58 measurements on simulated side impact collisions. The fatality (binary 0 – 1 random variable, 1 means the crash resulted in fatality) and three covariates (age of the driver, velocity of the automobile, maximal acceleration measured on the subject’s abdomen) were measured. Härdle and Stoker estimated  $\beta_0$  and fitted  $m$  in a non-parametric way and concluded that the link function is of "distribution

**Table 3**  $H_0$  (logistic regression function,  $\beta$  is 3-dimensional) is false

		BS	Stute/Zhu	Ecdffs (red: BS, blue: Stute/ Zhu)
$n = 50$	$\alpha = 0.05$	49.0%	44.2%	
	$\alpha = 0.01$	24.9%	21.3%	
$n = 100$	$\alpha = 0.05$	79.9%	78.3%	
	$\alpha = 0.01$	54.5%	53.6%	

**Table 4**  $H_0$  (logistic regression function,  $\beta$  is 3-dimensional) is false

		BS	Stute/Zhu	Ecdffs (red: BS, blue: Stute/ Zhu)
$n = 50$	$\alpha = 0.05$	6.0%	5.4%	
	$\alpha = 0.01$	1.4%	0.9%	
$n = 100$	$\alpha = 0.05$	4.3%	5.4%	
	$\alpha = 0.01$	0.8%	0.7%	

type", i.e., non-decreasing in  $\beta_0^T x$ , as in the logit or probit case. They did not check if a GLM would fit the data at all. We tested, if (after a standardization) a GLM with a logit or probit link function is appropriate for the data set. Based on the bootstrap approach the  $p$  value for a logit link function is 0.047, for a probit link function 0.049. Thus, in both cases, the model is rejected. Stute and Zhu (2002) also applied their approach to this data set and came to the same result.

## 4 Discussion

Our small simulation study indicates that the bootstrap approach has slightly better empirical power than the Stute & Zhu, (2002) approach. This is noteworthy because the Stute and Zhu approach was conducted here under an additional assumption (elliptically contoured distributions) that is unnecessary for the bootstrap approach. If this additional assumption is not fulfilled, then non-parametric regression estimation has to be applied in the Stute and Zhu procedure, but this entails further problems (choice of smoothing parameter) and can have negative effects on the power of the test. For the bootstrap method all these problems do not exist.

The resampling procedure guarantees that the bootstrap data are always generated under the null hypothesis, regardless of whether the original data satisfy the null hypothesis or not. Consequently, the distribution of a test statistic based on the bootstrap data fits the null hypothesis. If the test statistic based on the original data lies at the edge of this bootstrap-based distribution, then this indicates a violation of the null hypothesis. It is important to note that the sample size is also considered in the approximating distribution by the bootstrap approach. In the approximation with the asymptotic distribution this is not given in the last consequence. We assume that the slight improvement with respect to the empirical power is based on this. That the consideration of the sample size in the approximating distribution can be advantageous compared to the approximation by the limiting distribution, Singh, (1981) was able to prove for the classical bootstrap and the standardized mean. However, this is not studied further in our paper, but should be addressed theoretically in future work.

The bootstrap method is easier to implement because it is not as technically demanding as the method of Stute and Zhu. However, it is more complex in terms of computing time. The latter is always of great importance if the method is to be used on a large scale.

## 5 Proofs

**Proof of Lemma 1** Define  $\mathcal{F} = \{I(\beta^T \cdot \leq u)m(\beta^T \cdot)\bar{m}(\beta^T \cdot), \beta \in \mathbb{R}^d, u \in \mathbb{R}\}$ . Following Lemma 7,  $\mathcal{F}$  is GC. Thus (i) is true.

For (ii) check that

$$\begin{aligned} & \sup_{|\beta - \beta_0| \leq \varepsilon, u \in \mathbb{R}} \left| \mathbb{E}(I(\beta^T X \leq u)m(\beta^T X)\bar{m}(\beta^T X)) \right. \\ & \quad \left. - \mathbb{E}(I(\beta_0^T X \leq u)m(\beta_0^T X)\bar{m}(\beta_0^T X)) \right| \\ & \leq \sup_{|\beta - \beta_0| \leq \varepsilon, u \in \mathbb{R}} \mathbb{E} \left( \left| m(\beta^T X)\bar{m}(\beta^T X) - \mathbb{E}(m(\beta_0^T X)\bar{m}(\beta_0^T X)) \right| \right) \\ & \quad + \sup_{|\beta - \beta_0| \leq \varepsilon, u \in \mathbb{R}} \left| \mathbb{E}(I(\beta^T X \leq u)m(\beta_0^T X)\bar{m}(\beta_0^T X)) \right. \\ & \quad \left. - \mathbb{E}(I(\beta_0^T X \leq u)m(\beta_0^T X)\bar{m}(\beta_0^T X)) \right| \end{aligned}$$

Due to assumption (C1) the dominated convergence theorem yields that the first term converges to 0 as  $\varepsilon \rightarrow 0$ .

Denote the second term as  $\sup_{|\beta - \beta_0| \leq \varepsilon, u \in \mathbb{R}} A(\beta, u)$  and choose  $K > 0$  and check that

$$\begin{aligned} & A(\beta, u) \\ & \leq \mathbb{E} \left( m(\beta_0^T X)\bar{m}(\beta_0^T X) \left| I(\beta^T X \leq u) - I(\beta_0^T X \leq u) \right| I(|X| \leq K) \right) \\ & \quad + \mathbb{E} \left( m(\beta_0^T X)\bar{m}(\beta_0^T X) \left| I(\beta^T X \leq u) - I(\beta_0^T X \leq u) \right| I(|X| > K) \right) \\ & = A_1(\beta, \beta_0, u, K) + A_2(\beta, \beta_0, u, K) \end{aligned}$$

Select  $\gamma > 0$  to get

$$\begin{aligned} & A_1(\beta, \beta_0, u, K) \\ & \leq \mathbb{E} \left( m(\beta_0^T X)\bar{m}(\beta_0^T X) \left| I(\beta^T X \leq u) - I(\beta_0^T X \leq u) \right| \right. \\ & \quad \cdot I(|X| \leq K) I(|\beta^T X - \beta_0^T X| \leq \gamma) \\ & \quad \left. + I(|\beta - \beta_0| > \gamma/K) \mathbb{E}(m(\beta_0^T X)\bar{m}(\beta_0^T X)) \right) \\ & \leq \mathbb{E} \left( m(\beta_0^T X)\bar{m}(\beta_0^T X) \left| I(\beta_0^T X \leq u + \gamma) - I(\beta_0^T X \leq u) \right| \right) \\ & \quad + I(|\beta - \beta_0| > \gamma/K) \mathbb{E}(m(\beta_0^T X)\bar{m}(\beta_0^T X)) \\ & = H(u + \gamma, \beta_0) - H(u, \beta_0) \\ & \quad + I(|\beta - \beta_0| > \gamma/K) \mathbb{E}(m(\beta_0^T X)\bar{m}(\beta_0^T X)) \\ & = A_{1,1}(\beta_0, u, \gamma) + A_{1,2}(\beta, \beta_0, \gamma, K) \end{aligned}$$

Fix  $\delta > 0$ . Since  $\mathbb{E}(m(\beta_0^T X)\bar{m}(\beta_0^T X)) \leq 1$ , we can find a  $K > 0$  such that  $A_2(\beta, \beta_0, u, K) \leq \delta$ . Due to assumption (B1),  $H(\cdot, \beta_0)$  is uniformly continuous and therefore we can find a  $\gamma > 0$  such that  $A_{1,1} < \delta$  uniformly in  $u$ . Furthermore, we can choose  $\varepsilon$  such that  $\varepsilon < \min(\gamma, \gamma/K)$  which yields that  $A_{1,2}(\beta, \beta_0, \gamma, K) = 0$  and, therefore,  $A(\beta, \beta_0, u) < 2\delta$ . This proves part (ii).

Since  $\beta_n \rightarrow \beta_0$ , w.p. 1, (iii) follows directly from (i) and (ii). □

**Proof of Theorem 1** To prove the Theorem, we will use Theorem 13.5 of Billingsley (1999). We first show, that the fidis of  $R_n^*$  converge in distribution to the fidis of  $R_\infty$ . Obviously,  $R_n^*$  has independent zero-mean summands, since

$$\begin{aligned} & \mathbb{E}_n^* (I(\beta_n^T X_i^* \leq u) (\delta_i^* - m(\beta_n^T X_i^*))) \\ &= \mathbb{E}_n^* (I(\beta_n^T X_i \leq u) (\delta_i^* - m(\beta_n^T X_i))) \\ &= I(\beta_n^T X_i \leq u) (\mathbb{E}_n^* (\delta_i^*) - m(\beta_n^T X_i)) \\ &= I(\beta_n^T X_i \leq u) (m(\beta_n^T X_i) - m(\beta_n^T X_i)) \\ &= 0. \end{aligned}$$

For the covariance of  $R_n^*$  we get for  $u_1, u_2 \in \mathbb{R}$

$$\begin{aligned} \text{Cov}_n^*(R_n^*(u_1), R_n^*(u_2)) &= \mathbb{E}_n^*(R_n^*(u_1)R_n^*(u_2)) \\ &= \mathbb{E}_n^* \left( \frac{1}{n} \sum_{i=1}^n I(\beta_n^T X_i \leq u_1) (\delta_i^* - m(\beta_n^T X_i)) \right. \\ &\quad \cdot \left. \sum_{j=1}^n I(\beta_n^T X_j \leq u_2) (\delta_j^* - m(\beta_n^T X_j)) \right) \\ &= \frac{1}{n} \sum_{1 \leq i, j \leq n} I(\beta_n^T X_i \leq u_1) I(\beta_n^T X_j \leq u_2) \\ &\quad \cdot \mathbb{E}_n^* \left( (\delta_i^* - m(\beta_n^T X_i)) (\delta_j^* - m(\beta_n^T X_j)) \right), \end{aligned}$$

where  $\delta_i^*$  and  $\delta_j^*$  are iid. Thus, if  $i \neq j$ , the expectation in the last equation is 0. Therefore, the last equation equals

$$\frac{1}{n} \sum_{i=1}^n I(\beta_n^T X_i \leq u_1 \wedge u_2) \mathbb{E}_n^* ((\delta_i^* - m(\beta_n^T X_i))^2).$$

Here, the expectation equals the conditional covariance of a binomial distribution with success probability  $m(\beta_n^T X_i)$ . Thus, for the last equation we get

$$\frac{1}{n} \sum_{i=1}^n I(\beta_n^T X_i \leq u_1 \wedge u_2) m(\beta_n^T X_i) \bar{m}(\beta_n^T X_i).$$

Due to Lemma 1 (iii), this converges to  $\mathbb{E}(I(\beta_0^T X \leq u_1 \wedge u_2) m(\beta_0^T X) \bar{m}(\beta_0^T X))$  uniformly in  $u$  for  $n \rightarrow \infty$ , w. p. 1. Thus, w. p. 1, the covariance function of the process  $R_n^*(u)$  converges to

$$K(u_s, u_t) = \text{Cov}(R_\infty(u_s), R_\infty(u_t)) = \int m(\beta_0^T X) \bar{m}(\beta_0^T X) I(\beta_0^T X \leq u_s \wedge u_t) d\mathbb{P}.$$

Now let  $k \in \mathbb{N}$  and choose  $-\infty \leq u_1 < \dots < u_k \leq \infty$ . Following Cramér-Wold, see Theorem 7.7 of Billingsley, (1999), we have to show that, w.p. 1, for every  $a \in \mathbb{R}^k$ ,  $a \neq 0$

$$\sum_{j=1}^k a_j R_n^*(u_j) \rightarrow \mathcal{N}(0, a^T \Sigma a), \text{ for } n \rightarrow \infty$$

in distribution, with  $\Sigma = (\sigma_{s,t})_{1 \leq s,t \leq k}$  and  $\sigma_{s,t} = \text{Cov}(R_\infty(u_s), R_\infty(u_t)) = \mathbb{E}(R_\infty(u_s), R_\infty(u_t)) = K(u_s, u_t)$ .

Set

$$\begin{aligned} Z_n^* &= \sum_{j=1}^k a_j R_n^*(u_j) = n^{-1/2} \sum_{i=1}^n \left( (\delta_i^* - m(\beta_n^T X_i)) \sum_{j=1}^k a_j I(\beta_n^T X_i \leq u_j) \right) \\ &= \sum_{i=1}^n \xi_{i,n}^* A_{i,n}, \end{aligned}$$

where  $\xi_{i,n}^* = n^{-1/2}(\delta_i^* - m(\beta_n^T X_i))$  and  $A_{i,n} = \sum_{j=1}^k a_j I(\beta_n^T X_i \leq u_j)$ . Here,  $\xi_{1,n}^*, \dots, \xi_{n,n}^*$  are independent and centered, and  $A_{1,n}, \dots, A_{n,n}$  are deterministic in the bootstrap setup. To show the asymptotic normality of  $Z_n^*$ , we apply Theorem 1.9.3 of Serfling, (1980) and prove the Lindeberg condition,

$$\frac{1}{\text{Var}_n^*(Z_n^*)^2} \sum_{i=1}^n \int (\xi_{i,n}^* A_{i,n})^2 I\left(|\xi_{i,n}^* A_{i,n}| > \varepsilon \sqrt{\text{Var}_n^*(Z_n^*)}\right) d\mathbb{P}^* \rightarrow 0,$$

as  $n \rightarrow \infty$ , is true w.p. 1 for each  $\varepsilon > 0$ .

First, check that

$$\begin{aligned} \text{Var}_n^*(Z_n^*) &= \text{Var}_n^*\left(\sum_{j=1}^k a_j R_n^*(u_j)\right) \\ &= \sum_{1 \leq s,t \leq k} a_s \text{Cov}_n^*(R_n^*(u_s), R_n^*(u_t)) a_t \\ &\rightarrow \sum_{1 \leq s,t \leq k} a_s \text{Cov}(R_\infty(u_s), R_\infty(u_t)) a_t \\ &= \sum_{1 \leq s,t \leq k} a_s K(u_s, u_t) a_t \\ &= a^T \Sigma a, \text{ as } n \rightarrow \infty, \text{ w.p. } 1. \end{aligned}$$

Since  $\Sigma$  is positive semi-definite,  $a^T \Sigma a \geq 0$ . If  $a^T \Sigma a = 0$ , Tschebyscheff’s inequality guarantees that  $Z_n^* = o_{\mathbb{P}^*}(1)$  and thus, for  $n \rightarrow \infty$ ,

$$Z_n^* \rightarrow \mathcal{N}(0, a^T \Sigma a), \text{ w.p. } 1.$$

Now, assume that  $a^T \Sigma a > 0$ . Obviously  $|A_{i,n}| \leq \|a\|k$ . Hence, for each  $\varepsilon > 0$ ,



$$\begin{aligned}
 & I\left(\left|\xi_{i,n}^* A_{i,n}\right| > \varepsilon \sqrt{\text{Var}_n^*(Z_n^*)}\right) \\
 & \leq I\left(\left|\xi_{i,n}^*\right| \|a\| k > \varepsilon \sqrt{\text{Var}_n^*(Z_n^*)}\right) \\
 & = I\left(\left|\delta_i^* - m(\beta_n^T X_i)\right| \|a\| k > \sqrt{n} \varepsilon \sqrt{\text{Var}_n^*(Z_n^*)}\right) \\
 & \leq I\left(\|a\| k > \sqrt{n} \varepsilon \sqrt{\text{Var}_n^*(Z_n^*)}\right) \\
 & = 0, \text{ as } n \rightarrow \infty, \text{ w.p. } 1.
 \end{aligned}$$

Thus, the indicator of the Lindeberg condition equals 0 as  $n \rightarrow \infty$  and therefore the Lindeberg condition is fulfilled, the finite dimensional distributions converge to  $\mathcal{N}(0, a^T \Sigma a)$ . This is part (i) of Theorem 13.5 of Billingsley, (1999).

For the tightness we use a modification of this Theorem, see Corollary 1, where  $F$  also depends on  $n$ . For this we assume that our process is only defined on the interval  $[0, 1]$ . If this is not the case, we can use a transformation to receive such a process.

Check that for  $0 \leq u_1 \leq u \leq u_2 \leq 1$

$$\begin{aligned}
 & \mathbb{E}_n^* \left( [R_n^*(u) - R_n^*(u_1)]^2 [R_n^*(u_2) - R_n^*(u)]^2 \right) \\
 & = \frac{1}{n^2} \mathbb{E}_n^* \left( \left[ \sum_{i=1}^n I(\beta_n^T X_i \leq u) (\delta_i^* - m(\beta_n^T X_i^*)) \right. \right. \\
 & \quad \left. \left. - \sum_{i=1}^n I(\beta_n^T X_i \leq u_1) (\delta_i^* - m(\beta_n^T X_i^*)) \right]^2 \right. \\
 & \quad \cdot \left[ \sum_{i=1}^n I(\beta_n^T X_i \leq u_2) (\delta_i^* - m(\beta_n^T X_i^*)) \right. \\
 & \quad \left. \left. - \sum_{i=1}^n I(\beta_n^T X_i \leq u) (\delta_i^* - m(\beta_n^T X_i^*)) \right]^2 \right) \\
 & = \frac{1}{n^2} \mathbb{E}_n^* \left( \left[ \sum_{i=1}^n I(u_1 < \beta_n^T X_i \leq u) (\delta_i^* - m(\beta_n^T X_i^*)) \right]^2 \right. \\
 & \quad \cdot \left. \left[ \sum_{i=1}^n I(u < \beta_n^T X_i \leq u_2) (\delta_i^* - m(\beta_n^T X_i^*)) \right]^2 \right).
 \end{aligned}$$

Now set

$$\alpha_i = I(u_1 < \beta_n^T X_i \leq u) (\delta_i^* - m(\beta_n^T X_i^*))$$

and

$$\beta_i = I(u < \beta_n^T X_i \leq u_2) (\delta_i^* - m(\beta_n^T X_i^*)).$$

Use this and check that

$$\begin{aligned} & \frac{1}{n^2} \mathbb{E}_n^* \left( \left( \sum_{i=1}^n \alpha_i \right) \left( \sum_{j=1}^n \beta_j \right) \right) \\ &= \frac{1}{n^2} \sum_{1 \leq i, j, k, l \leq n} \mathbb{E}_n^* (\alpha_i \alpha_j \beta_k \beta_l) \\ &= \frac{1}{n^2} \left( \sum_{1 \leq i \neq j \leq n} (\mathbb{E}_n^* (\alpha_i \alpha_j \beta_j \beta_j) + 2\mathbb{E}_n^* (\alpha_i \alpha_j \beta_i \beta_j)) + \sum_{1 \leq i \leq n} \mathbb{E}_n^* (\alpha_i^2 \beta_i^2) \right) \\ &= \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} \mathbb{E}_n^* (\alpha_i^2) \mathbb{E}_n^* (\beta_j^2), \end{aligned}$$

where the last equality follows since the  $\alpha_i$  and  $\beta_i$  are independent and since either  $I(u_1 < \beta_n^T X_i \leq u)$  or  $I(u < \beta_n^T X_i \leq u_2)$  equals 0. Now, recall the definition of  $\alpha_i$  and  $\beta_i$  to get

$$\begin{aligned} & \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} \mathbb{E}_n^* (\alpha_i^2) \mathbb{E}_n^* (\beta_j^2) \\ &= \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} \mathbb{E}_n^* \left( I(u_1 < \beta_n^T X_i \leq u) (\delta_i^* - m(\beta_n^T X_i^*))^2 \right) \\ & \quad \cdot \mathbb{E}_n^* \left( I(u < \beta_n^T X_j \leq u_2) (\delta_j^* - m(\beta_n^T X_j^*))^2 \right) \\ &\leq \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} m(\beta_n^T X_i) \bar{m}(\beta_n^T X_i) I(u_1 < \beta_n^T X_i \leq u_2) \\ & \quad \cdot m(\beta_n^T X_j) \bar{m}(\beta_n^T X_j) I(u_1 < \beta_n^T X_j \leq u_2) \\ &\leq \frac{1}{n^2} \sum_{1 \leq i, j \leq n} m(\beta_n^T X_i) \bar{m}(\beta_n^T X_i) I(u_1 < \beta_n^T X_i \leq u_2) \\ & \quad \cdot m(\beta_n^T X_j) \bar{m}(\beta_n^T X_j) I(u_1 < \beta_n^T X_j \leq u_2) \\ &= \left( \frac{1}{n} \sum_{i=1}^n m(\beta_n^T X_i) \bar{m}(\beta_n^T X_i) I(u_1 < \beta_n^T X_i \leq u_2) \right)^2 \\ &= \left( \frac{1}{n} \sum_{i=1}^n m(\beta_n^T X_i) \bar{m}(\beta_n^T X_i) I(\beta_n^T X_i \leq u_2) \right. \\ & \quad \left. - \frac{1}{n} \sum_{i=1}^n m(\beta_n^T X_i) \bar{m}(\beta_n^T X_i) I(\beta_n^T X_i \leq u_1) \right)^2 \\ &=: (H_n(u_2) - H_n(u_1))^2. \end{aligned}$$

Since  $H_n(u) \rightarrow K(u, u)$ , w.p. 1, due to assumption (B1), a continuous, non-decreasing function  $H$  with  $\sup_{u \in \mathbb{R}} |H_n(u) - H(u)| \rightarrow 0$  exists. Therefore, following Corollary 1 the process  $R_n^*$  is tight. □

**Proof of Lemma 2** Following Cramér-Wold, see Theorem 7.7 of Billingsley (1999), due to (B2) we have to show that, w.p. 1, for every  $a \in \mathbb{R}^d, a \neq 0$ ,

$$Z_n^* = n^{-1/2} \sum_{i=1}^n a^T l(\beta_n^T X_i, \delta_i^*) \rightarrow \mathcal{N}(0, a^T L(\beta_0) a),$$

in distribution for  $n \rightarrow \infty$ . According to Serfling, (1980), Theorem 1.9.3, this follows from the Lindeberg condition,

$$\begin{aligned} \frac{1}{\text{Var}_n^*(Z_n^*)} \frac{1}{n} \sum_{i=1}^n \int a^T l(\beta_n^T X_i, \delta_i^*) l^T(\beta_n^T X_i, \delta_i^*) a \\ I\left(|n^{-1/2} a^T l(\beta_n^T X_i, \delta_i^*)| > \varepsilon \sqrt{\text{Var}_n^*(Z_n^*)}\right) d\mathbb{P}^* \\ \rightarrow 0, \text{ as } n \rightarrow \infty. \end{aligned}$$

Use (C2) to get

$$\begin{aligned} \text{Var}_n^*(Z_n^*) &= \frac{1}{n} \sum_{i=1}^n a^T \mathbb{E}_n^*(l(\beta_n^T X_i, \delta_i^*) l^T(\beta_n^T X_i, \delta_i^*)) a = a^T L_n^*(\beta_n) a \\ &\rightarrow a^T L(\beta_0) a, \end{aligned}$$

as  $n \rightarrow \infty$ , w.p.1. Furthermore,  $L(\beta_0)$  is positive definite and  $a \neq 0$ , thus,  $a^T L(\beta_0) a > 0$  and it suffices to show that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \int a^T l(\beta_n^T X_i, \delta_i^*) l^T(\beta_n^T X_i, \delta_i^*) a \\ \cdot I\left(|n^{-1/2} a^T l(\beta_n^T X_i, \delta_i^*)| > \varepsilon \sqrt{\text{Var}_n^*(Z_n^*)}\right) d\mathbb{P}^* \rightarrow 0, \text{ as } n \rightarrow \infty. \end{aligned}$$

The integral equals

$$n^{-1} \sum_{i=1}^n a^T l(\beta_n^T X_i, \delta_i) l^T(\beta_n^T X_i, \delta_i) a I\left(|n^{-1/2} a^T l(\beta_n^T X_i, \delta_i)| > \varepsilon \sqrt{\text{Var}_n(Z_n)}\right)$$

Since  $\text{Var}_n^* \rightarrow a^T L(\beta_0) a$  we get for the indicator

$$\begin{aligned}
 & I\left(|n^{-1/2}a^T l(\beta_n^T X_i, \delta_i)| > \varepsilon \sqrt{\text{Var}_n(Z_n)}\right) \\
 &= I\left(|a^T l(\beta_n^T X_i, \delta_i)| > \varepsilon n^{1/2} \sqrt{\text{Var}_n(Z_n)}\right) \\
 &< I\left(|a^T l(\beta_n^T X_i, \delta_i)| > \varepsilon i^{1/2} \sqrt{a^T L(\beta_0) a / 2}\right),
 \end{aligned}$$

for  $1 \leq i \leq n$  and  $n$  sufficiently large.

Due to assumption (E2),

$$\begin{aligned}
 & \sum_{i=1}^n \mathbb{P}\left(|a^T l(\beta_n^T X_i, \delta_i)|^2 > \varepsilon i\right) \\
 &= \sum_{i=1}^n \mathbb{P}\left(|a^T l(\beta_n^T X, \delta_i)|^2 > \varepsilon i\right) \\
 &= \sum_{i=1}^n \varepsilon^{-1} \int_{[\varepsilon(i-1), \varepsilon i]} \mathbb{P}\left(|a^T l(\beta_n^T X, \delta_i)|^2 > \varepsilon i\right) dx \\
 &\leq \sum_{i=1}^n \varepsilon^{-1} \int_{[\varepsilon(i-1), \varepsilon i]} \mathbb{P}\left(|a^T l(\beta_n^T X, \delta_i)|^2 > x\right) dx \\
 &= \varepsilon^{-1} \int_0^\infty \mathbb{P}\left(|a^T l(\beta_n^T X, \delta_i)|^2 > x\right) dx \\
 &= \varepsilon^{-1} \mathbb{E}\left(|a^T l(\beta_n^T X, \delta_i)|^2\right) \\
 &< \infty.
 \end{aligned}$$

Thus, Borel-Cantelli yields

$$\limsup_{i \rightarrow \infty} \frac{|a^T l(\beta_n^T X_i, \delta_i^*)|}{\sqrt{i}} = 0,$$

w.p. 1. Therefore, the indicator equals 0 as  $n \rightarrow \infty$ , and the Lindeberg condition is fulfilled. □

**Proof of Lemma 3** Since the half-spaces in  $\mathbb{R}^d$  are a GC-class and  $w_j(X, \beta)$  is integrable, see assumption (F2), Corollary 9.27 of Kosorok (2008) yields

$$\sup_{\beta \in \mathbb{R}^d, u \in \mathbb{R}} \left| n^{-1} \sum_{i=1}^n w_j(X_i, \beta_0) I(\beta^T X_i \leq u) - W_j(u, \beta) \right| \rightarrow 0, \text{ as } n \rightarrow \infty, \text{ w.p. 1.}$$

Due to assumption (A1), for every  $\varepsilon > 0$  we get

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \sup_{u \in \mathbb{R}} \left| n^{-1} \sum_{i=1}^n w_j(X_i, \beta_0) I(\beta_n^T X_i \leq u) - W_j(u, \beta_0) \right| \\ & \leq \limsup_{n \rightarrow \infty} \sup_{\beta \in \mathbb{R}^d, u \in \mathbb{R}} \left| n^{-1} \sum_{i=1}^n w_j(X_i, \beta_0) I(\beta^T X_i \leq u) - W_j(u, \beta) \right| \\ & \quad + \sup_{|\beta - \beta_0| < \varepsilon, u \in \mathbb{R}} \left| W_j(u, \beta) - W_j(u, \beta_0) \right| \\ & = \sup_{|\beta - \beta_0| < \varepsilon, u \in \mathbb{R}} \left| W_j(u, \beta) - W_j(u, \beta_0) \right|, \end{aligned}$$

w.p. 1. Furthermore, the last term on the right side tends to 0 as  $\varepsilon \rightarrow 0$ . This is part (i).

For part (ii) we get

$$\begin{aligned} & \sup_{u \in \mathbb{R}} \left| n^{-1} \sum_{i=1}^n (w_j(X_i, \hat{\beta}_n^*(X_i)) - w_j(X_i, \beta_0)) I(\beta_n^T X_i \leq u) \right| \\ & \leq n^{-1} \sum_{i=1}^n \sup_{|\beta - \beta_0| < \varepsilon} \left| w_j(X_i, \beta) - w_j(X_i, \beta_0) \right| + o_{\mathbb{P}_n^*}(1), \end{aligned}$$

since, due to (A1) and Lemma 2, for  $\varepsilon > 0$ ,  $\mathbb{P}_n^*(|\beta_n^* - \beta_0| > \varepsilon) \rightarrow 0$  as  $n \rightarrow \infty$ , w.p. 1.

Furthermore, as  $n \rightarrow \infty$

$$\begin{aligned} & n^{-1} \sum_{i=1}^n \sup_{|\beta - \beta_0| < \varepsilon} \left| w_j(X_i, \beta) - w_j(X_i, \beta_0) \right| \\ & \rightarrow \mathbb{E} \left( \sup_{|\beta - \beta_0| < \varepsilon} \left| w_j(X_i, \beta) - w_j(X_i, \beta_0) \right| \right). \end{aligned}$$

Due to assumption (D2) and (E2), applying the dominated convergence theorem yields that the expectation on the right side tends to 0 as  $\varepsilon \rightarrow 0$ . □

**Proof of Theorem 2** Check that

$$\begin{aligned} R_n^{1*}(u) &= n^{-1/2} \sum_{i=1}^n (\delta_i^* - m(\beta_n^{*T} X_i)) I(\beta_n^T X_i \leq u) \\ &= n^{-1/2} \sum_{i=1}^n (\delta_i^* - m(\beta_n^T X_i)) I(\beta_n^T X_i \leq u) \\ & \quad - n^{-1/2} \sum_{i=1}^n (m(\beta_n^{*T} X_i) - m(\beta_n^T X_i)) I(\beta_n^T X_i \leq u) \\ &= R_n^*(u) - S_n^*(u). \end{aligned}$$

Since we already dealt with  $R_n^*(u)$  in Theorem 1, we now have to handle  $S_n^*(u)$ . It follows from assumptions (A1) and Lemma 2 that

$$\mathbb{P}_n^*(|\beta_n^* - \beta_0| > \varepsilon) \rightarrow 0, \text{ as } n \rightarrow \infty, \text{ w.p. } 1,$$

for  $\varepsilon > 0$ . Thus, we can assume that  $\beta_n^*$  and  $\beta_n$  are in the neighborhood of  $\beta_0$ . Following assumption (D2) we can apply Taylor’s expansion to get

$$m(\beta_n^{*T} x) = m(\beta_n^T x) + (\beta_n^* - \beta_n)^T w(x, \hat{\beta}_n^*(x)),$$

where  $\hat{\beta}_n^*(x)$  is in the line segment connecting  $\beta_n^*$  and  $\beta_n$ . Thus we can write  $S_n^*(u)$  as follows:

$$\begin{aligned} S_n^*(u) &= n^{1/2}(\beta_n^* - \beta_n)^T n^{-1} \sum_{i=1}^n w(X_i, \hat{\beta}_n^*(X_i))I(\beta_n^{*T} X_i \leq u) + o_{\mathbb{P}_n^*}(1) \\ &= n^{1/2}(\beta_n^* - \beta_n)^T W(u, \beta_0) \\ &\quad + n^{1/2}(\beta_n^* - \beta_n)^T n^{-1} \sum_{i=1}^n (w(X_i, \hat{\beta}_n^*(X_i)) - w(X_i, \beta_0))I(\beta_n^T X_i \leq u) \\ &\quad + n^{1/2}(\beta_n^* - \beta_n)^T \left( n^{-1} \sum_{i=1}^n w(X_i, \beta_0)I(\beta_n^T X_i \leq u) - W(u, \beta_0) \right) \\ &\quad + o_{\mathbb{P}_n^*}(1). \end{aligned}$$

Lemma 3 now yields that

$$S_n^*(u) = n^{1/2}(\beta_n^* - \beta_n)^T W(u, \beta_0) + o_{\mathbb{P}_n^*}(1), \text{ w.p. } 1,$$

and with (B2) and (E2)

$$S_n^*(u) = n^{-1/2} \sum_{i=1}^n l(\beta_n^T X_i, \delta_i^*)W(u, \beta_0) + o_{\mathbb{P}_n^*}(1), \text{ w.p. } 1,$$

uniformly in  $u$ .

Now define

$$\begin{aligned} \hat{R}_n^{1*}(u) &= n^{-1/2} \sum_{i=1}^n ((\delta_i^* - m(\beta_n^T X_i))I(\beta_n^T X_i \leq u) - l^T(\beta_n^T X_i, \delta_i^*)W(u, \beta_0)) \\ &= R_n^*(u) - n^{-1/2} \sum_{i=1}^n l^T(\beta_n^T X_i, \delta_i^*)W(u, \beta_0), \end{aligned}$$

which is asymptotically equivalent to  $R_n^{1*}$ , see Theorem 4.1 of Billingsley (1999). Furthermore, following the proof of Theorem 1,  $R_n^*(u)$  is tight in  $D[-\infty, \infty]$  and due to Lemma 2  $n^{-1/2} \sum_{i=1}^n l^T(\beta_n^T X_i, \delta_i^*)$  converges to a zero mean multivariate normal distribution with covariance matrix  $L(\beta_0)$ , w. p. 1.

Furthermore, assumption (F2) yields that  $W(\cdot)$  is continuous.

Thus,  $n^{-1/2} \sum_{i=1}^n l^T(\beta_n^T X_i, \delta_i^*)W(u, \beta_0)$  is tight in  $C[-\infty, \infty]$  and therefore also tight in  $D[-\infty, \infty]$ . Finally, w.p. 1,  $\hat{R}_n^{1*}(u)$  is tight in  $D[-\infty, \infty]$ .

Now let  $k \in \mathbb{N}$  and choose  $-\infty \leq u_1 < \dots < u_k \leq \infty$ . Following Cramér-Wold, see Theorem 7.7 of Billingsley, (1999), we have to show that, w.p. 1, for every  $a \in \mathbb{R}^k$ ,  $a \neq 0$

$$Z_n^{1*} = \sum_{j=1}^k a_j \hat{K}_n^{1*}(u_j) \rightarrow \mathcal{N}(0, a^T \Sigma a), \text{ for } n \rightarrow \infty,$$

in distribution, with  $\Sigma = (\sigma_{s,t})_{1 \leq s,t \leq k}$  and  $\sigma_{s,t} = Cov(R_\infty^1(u_s), R_\infty^1(u_t)) = \mathbb{E}(R_\infty^1(u_s), R_\infty^1(u_t)) = \hat{K}(u_s, u_t)$ .

We can rearrange the terms to

$$\begin{aligned} Z_n^{1*} &= \sum_{i=1}^n \left( \frac{\delta_i^* - m(\beta_n^T X_i)}{\sqrt{n}} \sum_{j=1}^k a_j I(\beta_n^T X_i < u_j) \right. \\ &\quad \left. - \frac{l^T(\beta_n^T X_i, \delta_i^*)}{\sqrt{n}} \sum_{j=1}^k a_j W(u_j, \beta_0) \right) \\ &= \sum_{i=1}^n \xi_{i,n}^* A_{i,n} - \eta_{i,n}^{*T} B, \end{aligned}$$

with  $\xi_{i,n}^* = n^{-1/2}(\delta_i^* - m(\beta_n^T X_i))$  and  $\eta_{i,n}^* = n^{-1/2}l(\beta_n^T X_i, \delta_i^*)$ . Obviously, those variables are centered and  $(\xi_{1,n}^*, \eta_{1,n}^*), \dots, (\xi_{n,n}^*, \eta_{n,n}^*)$  are independent. Additionally,  $A_{i,n}$  and  $B$  are deterministic with respect to  $\mathbb{P}_n^*$ . Thus, we get for the variance of  $Z_n^{1*}$

$$\begin{aligned} Var_n^*(Z_n^{1*}) &= \sum_{i=1}^n A_{i,n}^2 Var_n^*(\xi_{i,n}^*) + \sum_{i=1}^n B^T \mathbb{E}_n^*(\eta_{i,n}^* \eta_{i,n}^{*T}) B \\ &\quad - 2B^T \sum_{i=1}^n \mathbb{E}_n^*(\xi_{i,n}^* \eta_{i,n}^*) A_{i,n}. \end{aligned}$$

In the proof of Theorem 1 we have shown that

$$\sum_{i=1}^n A_{i,n}^2 Var_n^*(\xi_{i,n}^*) \rightarrow \sum_{1 \leq s,t \leq n} a_s K(u_s, u_t) a_t, \text{ as } n \rightarrow \infty, \text{ w.p. 1.}$$

Furthermore, due to assumption (C2)

$$\sum_{i=1}^n B^T \mathbb{E}_n^*(\eta_{i,n}^* \eta_{i,n}^{*T}) B \rightarrow B^T L(\beta_0) B = \sum_{1 \leq s,t \leq n} a_s W^T(u_s, \beta_0) L(\beta_0) W(u_t, \beta_0) a_t,$$

as  $n \rightarrow \infty$ , w.p. 1.

Now check that

$$\begin{aligned}
 & \mathbb{E}_n^*(\xi_{i,n}^* \eta_{i,n}^*) \\
 &= \frac{1}{n} \mathbb{E}_n^* \left( (\delta_i^* - m(\beta_n^T X_i)) l(\beta_n^T X_i, \delta_i^*) \right) \\
 &= \frac{1}{n} \mathbb{E}_n^* \left( (\delta_i^* - m(\beta_n^T X_i)) \left( \delta_i^* \frac{w(X_i, \beta_n)}{m(\beta_n^T X_i)} - (1 - \delta_i^*) \frac{w(X_i, \beta_n)}{1 - m(\beta_n^T X_i)} \right) \right) \\
 &= \frac{1}{n} \mathbb{E}_n^* \left( (\delta_i^* - m(\beta_n^T X_i)) \frac{(\delta_i^* - m(\beta_n^T X_i)) w(X_i, \beta_n)}{m(\beta_n^T X_i)(1 - m(\beta_n^T X_i))} \right) \\
 &= \frac{1}{n} w(X_i, \beta_n).
 \end{aligned}$$

Thus, for the last term, due to Lemma 2 (i), we get as  $n \rightarrow \infty$

$$\begin{aligned}
 & B^T \sum_{i=1}^n \mathbb{E}_n^*(\xi_{i,n}^* \eta_{i,n}^*) A_{i,n} \\
 &= \left( \sum_{l=1}^k a_l W(u_l, \beta_0) \right)^T \frac{1}{n} \sum_{i=1}^n \left( w(X_i, \beta_n) \sum_{j=1}^k a_j I(\beta_n^T X_i < u_j) \right) \\
 &= \sum_{1 \leq s, t \leq k} \left[ a_s W(u_s, \beta_0)^T \frac{1}{n} \sum_{i=1}^n (w(X_i, \beta_n) I(\beta_n^T X_i < u_t)) a_t \right] \\
 &\rightarrow \sum_{1 \leq s, t \leq k} a_s W(u_s, \beta_0)^T W(u_t, \beta_0) a_t.
 \end{aligned}$$

And finally, as  $n \rightarrow \infty$ , w.p. 1,

$$\text{Var}_n^*(Z_n^{1*}) \rightarrow \sum_{1 \leq s, t \leq n} a_s \hat{K}(u_s, u_t) a_t = a^T \Sigma a.$$

Assume that  $a^T \Sigma a > 0$ . Then we have to prove the Lindeberg condition

$$\begin{aligned}
 & \frac{1}{\text{Var}_n^*(Z_n^{1*})} \sum_{i=1}^n \int (\xi_{i,n}^* A_{i,n} - \eta_{i,n}^{*T} B)^2 \\
 & \quad \cdot I \left( \left| \xi_{i,n}^* A_{i,n} - \eta_{i,n}^{*T} B \right| > \varepsilon \sqrt{\text{Var}_n^*(Z_n^{1*})} \right) d\mathbb{P}^* \rightarrow 0,
 \end{aligned}$$

as  $n \rightarrow \infty$ , w.p. 1.

The integral equals

$$n^{-1} \sum_{i=1}^n (\xi_{i,n}^* A_{i,n} - \eta_{i,n}^{*T} B)^2 I \left( \left| \xi_{i,n}^* A_{i,n} - \eta_{i,n}^{*T} B \right| > \varepsilon \sqrt{\text{Var}_n^*(Z_n^{1*})} \right).$$

Since  $\text{Var}_n^*(Z_n^{1*}) \rightarrow a^T L(\beta_0) a$  we get for the indicator



$$\begin{aligned}
 & I\left(\left|\xi_{i,n}A_{i,n} - \eta_{i,n}^T B\right| > \varepsilon \sqrt{\text{Var}_n(Z_n^1)}\right) \\
 &= I\left(\left|(\delta_i - m(\beta_n^T X_i))A_{i,n} - n^{-1/2}l^T(\beta_n^T X_i, \delta_i)B\right| > \varepsilon n^{1/2} \sqrt{\text{Var}_n(Z_n^1)}\right) \\
 &< I\left(\left|(\delta_i - m(\beta_n^T X_i))A_{i,n} - n^{-1/2}l^T(\beta_n^T X_i, \delta_i)B\right| > \varepsilon i^{1/2} \sqrt{a^T L(\beta_0)a/2}\right),
 \end{aligned}$$

for  $1 \leq i \leq n$  an  $n$  sufficiently large.

Since  $|(\delta_i - m(\beta_n^T X_i))A_{i,n}|$  and  $B$  are bounded, assumption (E2) yields

$$\begin{aligned}
 & \sum_{i \geq 1} \mathbb{P}\left(\left|(\delta_i - m(\beta_n^T X_i))A_{i,n} - n^{-1/2}l^T(\beta_n^T X_i, \delta_i)B\right|^2 > \varepsilon i\right) \\
 &= \sum_{i \geq 1} \mathbb{P}\left(\left|(\delta_i - m(\beta_n^T X_i))A_{i,n} - n^{-1/2}l^T(\beta_n^T X_i, \delta_i)B\right|^2 > \varepsilon i\right) \\
 &= \sum_{i \geq 1} \varepsilon^{-1} \int_{[\varepsilon(i-1), \varepsilon i]} \mathbb{P}\left(\left|(\delta_i - m(\beta_n^T X))A_{i,n} - n^{-1/2}l^T(\beta_n^T X, \delta_i)B\right|^2 > \varepsilon i\right) dx \\
 &\leq \sum_{i \geq 1} \varepsilon^{-1} \int_{[\varepsilon(i-1), \varepsilon i]} \mathbb{P}\left(\left|(\delta_i - m(\beta_n^T X))A_{i,n} - n^{-1/2}l^T(\beta_n^T X, \delta_i^*)B\right|^2 > x\right) dx \\
 &= \varepsilon^{-1} \int_0^\infty \mathbb{P}\left(\left|(\delta_i - m(\beta_n^T X))A_{i,n} - n^{-1/2}l^T(\beta_n^T X, \delta_i)B\right|^2 > x\right) dx \\
 &= \varepsilon^{-1} \mathbb{E}\left(\left|(\delta_i - m(\beta_n^T X))A_{i,n} - n^{-1/2}l^T(\beta_n^T X, \delta_i)B\right|^2\right) \\
 &\leq \varepsilon^{-1} \left(\mathbb{E}\left(\left|(\delta_i - m(\beta_n^T X))A_{i,n}\right|^2\right) + 2\mathbb{E}\left(\left|(\delta_i - m(\beta_n^T X))A_{i,n}\right| \left|n^{-1/2}l^T(\beta_n^T X, \delta_i)B\right|\right) + \mathbb{E}\left(\left|n^{-1/2}l^T(\beta_n^T X, \delta_i)B\right|^2\right)\right) \\
 &< \infty.
 \end{aligned}$$

Thus, Borel-Cantelli yields

$$\limsup_{i \rightarrow \infty} \frac{\left|(\delta_i - m(\beta_n^T X_i))A_{i,n} - n^{-1/2}l^T(\beta_n^T X_i, \delta_i)B\right|}{\sqrt{i}} = 0,$$

w.p. 1. Therefore, the indicator equals 0 as  $n \rightarrow \infty$ , and the Lindeberg condition is fulfilled and the finite dimensional distributions converge against a centered normal distribution with variance  $a^T \Sigma a$  in distribution as  $n \rightarrow \infty$ . □

### Appendix

The following results are used in the proofs in Section 5.

**Lemma 4**  $\tilde{\mathcal{H}} := \{I_H | H \in \mathcal{H}\}$ , with  $\mathcal{H} = \{\{x | \beta^T x \leq u\}, \beta \in \mathbb{R}^d, u \in \mathbb{R}\}$  is a Vapnik-Cervonenkis-class (VC), has a bounded uniform entropy integral (BUEI) with envelope  $\tilde{H} = 1$ , and is pointwise measurable (PM).

**Proof** Due to Lemma 9.12 (i) of Kosorok (2008), the half-spaces built a VC-class. Thus,  $\tilde{\mathcal{H}}$  is also a VC-class and therefore also BUEI, see Lemma 9.8 and Theorem 9.3 of Kosorok (2008). Since all values of  $\tilde{\mathcal{H}}$  are smaller or equal 1,  $\tilde{\mathcal{H}}$  has the envelope  $\tilde{H} = 1$ . Following Kosorok, (2008),  $\tilde{\mathcal{H}}$  is also PM. □

**Lemma 5**  $\mathcal{G} := \{g : g(x) = \beta^T x, \beta \in \mathbb{R}^d\}$  is a VC-class, BUEI and PM.

**Proof** Since  $\mathcal{G}$  is a finite dimensional vector space of measurable functions,  $\mathcal{G}$  is a VC-class and therefore also BUEI, see Lemma 9.6 and Theorem 9.3 of Kosorok (2008). Furthermore, we can choose the subset  $\mathcal{G}_\mathbb{Q}$  of  $\mathcal{G}$  such that the  $\beta$  are in the rational subset of  $\mathbb{R}^d$ , which is countable, and therefore  $\mathcal{G}_\mathbb{Q}$  is also countable. Obviously, for each  $f \in \mathcal{G}$  we can find a sequence  $\{g_m\} \in \mathcal{G}_\mathbb{Q}$  such that  $g_m(x) \rightarrow f(x)$  for each  $x \in \mathcal{X}$ . Thus  $\mathcal{G}$  is PM. □

**Lemma 6** Assume  $m$  is continuous and  $m'$  is bounded.

Then  $\tilde{\mathcal{G}} := \{m(g)\bar{m}(g) | g \in \mathcal{G}\}$  is BUEI with envelope  $\tilde{G} = 1$  and PM.

**Proof** Define  $\phi(t) := m(t)\bar{m}(t) = m(t) - m^2(t)$  and use Taylor expansion to get that

$$\begin{aligned} & |\phi(\beta_1^T x) - \phi(\beta_2^T x)|^2 \\ &= \left| \phi'(t^*(\beta_1, \beta_2, x)) (\beta_1^T x - \beta_2^T x) \right|^2 \\ &= \left| (m'(t^*(\beta_1, \beta_2, x)) - 2m'(t^*(\beta_1, \beta_2, x))m(t^*(\beta_1, \beta_2, x))) (\beta_1^T x - \beta_2^T x) \right|^2 \\ &= \left| m'(t^*(\beta_1, \beta_2, x)) (1 - 2m(t^*(\beta_1, \beta_2, x))) \right|^2 \left| \beta_1^T x - \beta_2^T x \right|^2 \\ &\leq \left| m'(t^*(\beta_1, \beta_2, x)) \right|^2 \left| \beta_1^T x - \beta_2^T x \right|^2, \end{aligned}$$

where  $t^*(\beta_1, \beta_2, x)$  is between  $\beta_1^T x$  and  $\beta_2^T x$ . The last inequality is correct since  $0 \leq m \leq 1$ . Furthermore,  $m'$  is bounded and due to Lemma 5 we can apply Lemma

9.13 of Kosorok, (2008) to get that  $\tilde{\mathcal{G}}$  is BUEI. Additionally,  $0 \leq m\bar{m} \leq 1$  and therefore  $\tilde{\mathcal{G}}$  has the envelope  $\tilde{G} = 1$ . Also,  $m$  is continuous and  $\mathcal{G}$  is PM, thus  $\tilde{\mathcal{G}}$  is also PM, see Lemma 8.10 of Kosorok, (2008).  $\square$

**Lemma 7** Assume  $m$  is continuous and  $m'$  is bounded. Combine  $\tilde{\mathcal{G}}$  and  $\tilde{\mathcal{H}}$  to get  $\mathcal{F} := \{m(g)\bar{m}(g)I_H, g \in \mathcal{G}, I_H \in \tilde{\mathcal{H}}\}$ .  $\mathcal{F}$  is BUEI with envelope  $F = 1$ , PM, Donsker and GC.

**Proof** Since  $\tilde{\mathcal{G}}$  and  $\tilde{\mathcal{H}}$  are BUEI with envelopes  $\tilde{G}$  and  $\tilde{H}$  and PM,  $\mathcal{F} = \tilde{\mathcal{G}}\tilde{\mathcal{H}}$  is BUEI with envelope  $F = \tilde{G}\tilde{H} = 1$  and PM, see (Kosorok, 2008, Lemma 9.17(v)).

Furthermore,  $E(F^2) < \infty$ . Following Kosorok (2008, page 165) Kosorok (2008),  $\mathcal{F}$  is Donsker and therefore also a GC.  $\square$

**Corollary 1** Assume that  $Y$  is a process in  $D([0, 1])$  and that w.p. 1, as  $n \rightarrow \infty$

$$(Y(t_1)^n, \dots, Y(t_k)^n) \rightarrow (Y(t_1), \dots, Y(t_k))$$

in distribution for points  $t_i$  of  $[0, 1]$ , that w.p. 1

$$Y(1) - Y(1 - \delta) \rightarrow 0, \delta \rightarrow 0,$$

in distribution, and that, for  $r \leq s \leq t, n \geq 1$  and  $\lambda > 0$ ,

$$\mathbb{P}(|Y(s)^n - Y(r)^n|^{2\beta} \wedge |Y(t)^n - Y(s)^n|^{2\beta} \geq \lambda) \leq \frac{1}{\lambda^{4\beta}} (H_n(t) - H_n(r))^{2\alpha}, \quad (1)$$

where  $\beta \geq 0$  and  $\alpha > 1/2$  and there exists  $H$ , a continuous, non-decreasing function on  $[0, 1]$  with  $\sup_{s \in [0,1]} |H_n(s) - H(s)| \rightarrow 0$ . Then  $Y^n \rightarrow Y$  as  $n \rightarrow \infty$  in distribution w.p.1.

(1) follows from:

$$\mathbb{E}(|Y(s)^n - Y(r)^n|^{2\beta} |Y(t)^n - Y(s)^n|^{2\beta} \geq \lambda) \leq (H_n(t) - H_n(r))^{2\alpha}.$$

**Proof** Following Theorem 13.5 of Billingsley, (1999), it is sufficient to show, that for  $\epsilon > 0, \eta > 0$ , there exists a  $\delta$  with  $0 < \delta < 1$  and a  $n_0$  such that  $\mathbb{P}_n[y : \omega''_n(\delta) \geq \epsilon] \leq \eta, n \geq n_0$ , where  $\omega''_n$  is the modulus of continuity. Apply Theorem 10.4 (Billingsley 1999) with  $Y^n$  in the role of  $\gamma$ . Then (10.20) is the same as (1). Let  $T = [0, 1]$ . Thus we get by (10.2)

$$\begin{aligned}
\mathbb{P}_n(y : \omega_n''(\delta) \geq \epsilon) &\leq \frac{2K}{\epsilon^{4\beta}} \sup_{0 \leq t \leq 1-2\delta} (H_n(t+2\delta) - H_n(t))^{2\alpha-1} \\
&\leq \frac{2K}{\epsilon^{4\beta}} \left( \sup_{0 \leq t \leq 1-2\delta} |H_n(t+2\delta) - H(t+2\delta)|^{2\alpha-1} \right. \\
&\quad + \sup_{0 \leq t \leq 1-2\delta} |H(t) - H_n(t)|^{2\alpha-1} \\
&\quad \left. + \sup_{0 \leq t \leq 1-2\delta} |H(t+2\delta) - H(t)|^{2\alpha-1} \right) \\
&\leq \frac{\eta}{3} + \frac{\eta}{3} + \frac{\eta}{3}.
\end{aligned}$$

This follows from Glivenko-Cantelli (first 2 summands) and the continuity of  $H$  (last summand). Thus, for given  $\epsilon$  and  $\mu$  we can choose  $\delta$  such that the right site of the inequality is less than  $\eta$ .  $\square$

**Acknowledgements** We thank Cornelia Krome for her helpful notes and her careful reading of the manuscript. Furthermore, we would like to thank Professor Li-Xing Zhu for providing us with the source code of further simulation studies of their method.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Data availability** The data used in the Real Data Application can be found in the paper of Härdle and Stoker (1989).

**Declarations**

**Conflict of interest** Not applicable

**Code availability** The code used in Section 3 can be provided.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Agresti, A. (2002). *Categorical data analysis, second edn.* Wiley Series in Probability and Statistics. New York: Wiley-Interscience [John Wiley & Sons].
- Bass, R. F. (2011). *Stochastic processes.* Cambridge series in statistical and probabilistic mathematics. Cambridge University Press.

- Billingsley, P. (1999). *Convergence of probability measures. second edn. Wiley series in probability and statistics: probability and statistics*. New York: John Wiley & Sons Inc.
- Dikta, G., Kvesic, M., & Schmidt, C. (2006). Bootstrap approximations in model checks for binary data. *Journal of the American Statistical Association*, 101(474), 521–530.
- Härdle, W., & Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, 84(408), 986–995.
- Kosorok, M. (2008). *Introduction to empirical processes and semiparametric inference*. New York: Springer.
- Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427), 846–866.
- Serfling, R. (1980). *Approximation theorems of mathematical statistics. [nachdr.] edn. Wiley series in probability and mathematical statistics : probability and mathematical statistics*. NY: Wiley.
- Singh, K. (1981). On the Asymptotic Accuracy of Efron's Bootstrap. *The Annals of Statistics*, 9(6), 1187–1195.
- Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*, 25(2), 613–641.
- Stute, W., & Zhu, L. X. (2002). Model checks for generalized linear models. *Scandinavian Journal of Statistics*, 29(3), 535–545.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.