Annals of the Institute of Statistical Mathematics (2024) 76:159–183
https://doi.org/10.1007/s10463-023-00880-8

Comparing regression curves: an L^1-point of view

Patrick Bastian[1] · Holger Dette[1] · Lukas Koletzko[1] · Kathrin Möllenhoff[2]

Received: 2 February 2023 / Revised: 25 June 2023 / Accepted: 3 July 2023 /
Published online: 30 August 2023

Abstract

In this paper, we compare two regression curves by measuring their difference by the area between the two curves, represented by their L^1-distance. We develop asymptotic confidence intervals for this measure and statistical tests to investigate the similarity/equivalence of the two curves. Bootstrap methodology specifically designed for equivalence testing is developed to obtain procedures with good finite sample properties and its consistency is rigorously proved. The finite sample properties are investigated by means of a small simulation study.

Keywords Equivalence testing · Comparison of curves · Bootstrap · Directional Hadamard differentiability

1 Introduction

The comparison of regression curves is a common problem in applied regression analysis. Usually these curves correspond to the means of a control and a treatment outcome where the predictor variable is an adjustable parameter, such as the time or a dose level, and an important question is whether the difference between the two curves is practically irrelevant. Borrowing ideas from testing for bioequivalence in

✉ Holger Dette
holger.dette@ruhr-uni-bochum.de

Patrick Bastian
patrick.bastian@ruhr-uni-bochum.de

Lukas Koletzko
lukas.koletzko@ruhr-uni-bochum.de

Kathrin Möllenhoff
kathrin.moellenhoff@hhu.de

[1] Fakultät für Mathematik, Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum, Germany

[2] Mathematisch-Naturwissenschaftliche Fakultät, Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany

population pharmacokinetics (see, for example, Chow and Liu 1992; Hauschke et al. 2007) numerous authors have addressed the problem of establishing the similarity between two regression functions by testing hypotheses of the form

$$H_0 \, : \, d \geq \epsilon \text{ versus } H_1 \, : \, d < \epsilon, \tag{1}$$

where d is a distance between the curves (which vanishes, if they are identical) and ϵ is a threshold, which defines when the difference between the two curves is considered as practically irrelevant.

Hypotheses of this type have found considerable interest in the literature. In many applications, the sample sizes are small such that nonparametric approaches are prohibitive and the relations between predictor and response for the two groups are modelled by nonlinear regression models with low-dimensional parameters. Liu et al. (2009) proposed tests for comparing linear models, while Gsteiger et al. (2011) suggested a bootstrap test for nonlinear models. The tests in these papers are based on the intersection–union principle (see, for example, Berger 1982) and a confidence band for the difference of the two regression models (see also Liu et al. 2007). Dette et al. (2018) pointed out that, by construction, these tests are very conservative and proposed an alternative approach, which has been successfully applied by Möllenhoff et al. (2022), among others. A common feature of the literature in this field consists in the fact that all methods use the maximal deviation $d_\infty := \sup_x |m_1(x) - m_2(x)|$ for the comparison of the curves, say m_1 and m_2 (here x denotes the predictor). While this metric has some attractive features such as good interpretability and a simple view of large distances, it is often too restrictive as one is interested in a "worst-case" scenario on a local scale. More precisely, if one uses the metric of maximum deviation between the two curves one is not able to decide for similarity if the curves are very similar for most points $x \in \mathcal{X}$, except for a "small" region. In such cases, an "average"

$$d_1 := \int |m_1(x) - m_2(x)| \mathrm{d}x$$

measuring the area between the curves might have advantages, and in this paper, we develop statistical inference tools if two regression curves are compared on the basis of this L^1-distance.

Our interest for this distance stems from the fact that the area under the curve (AUC) is a quite popular measure in biostatistics. For example, according to current guidelines by regulation authorities in both the US and the EU bioequivalence between a reference and a test product is to be assessed based on the comparison of their respective area under the time–concentration curves and their maximal concentrations (see US Food and Drug Administration 2003; EMA 2014 for details). For analyses of clinical trials with a time-to-event outcome, Royston and Parmar (2013) and McCaw et al. (2019) proposed the restricted mean survival time (RMST) as a possible alternative tool to the commonly used hazard ratio to estimate the treatment effect. By introducing this measure, Royston and Parmar (2013) and McCaw et al. (2019) addressed the well-known issue of assuming proportional hazards between different arms of the trial—an assumption, which is often unrealistic or obviously violated, but, however, only rarely assessed in practice (see Jachno et al. 2019 for a recent overview). The RMST, which is the area under the survival curve up to a specific time point, comes along without any assumption on the shape of the hazard ratio by calculating a treatment effect as the difference in RMST. We also refer to Cox and Czanner (2016) who investigated the L^1-distance for comparing survival distributions. Moreover, an important performance metric for assessing how well clinical risk prediction models distinguish between patients with and without a health outcome of interest is the area under receiver operating characteristic (ROC) curves (short AUROC), a tool of huge practical interest (see Pepe et al. 2013; Heller et al. 2016, among others). Also apart from medical questions, the AUROC is a common used tool in machine learning, arising whenever classifiers are evaluated or compared to each other regarding their power, see, for example, Bradley (1997).

As pointed out by Cox and Czanner (2016), the choice of L^1-distance for the comparison of curves poses several mathematical challenges, which are caused by the fact that the mapping $f \to \int_{\mathcal{X}} |f(x)|\mathrm{d}x$ is in general not (Hadamard-)differentiable. Cox and Czanner (2016) considered the distance $\int |S_1 f_2(x) - S_2 f_1(x)|\mathrm{d}x$ and restricted themselves to the situation where $S_1 f_2 \geq S_2 f_1$ to solve the differentiability problem. (In this case, the absolute value in the integral can be omitted.) In the context of testing for the equivalence of multinomial distributions, Ostrovski (2017) proposed to use a smooth approximation of the L^1-norm to avoid the differentiability problem.

Our approach for investigating the similarity between the regression functions m_1 and m_2 avoids such approximations and does not require that one regression function is larger than the other one. It is based on a direct estimate of the distance d_1 whose asymptotic properties are investigated in Sect. 2. The results are used for the construction of asymptotic confidence intervals for d_1 and a corresponding test for the hypotheses (1) by duality principles. In order to obtain less conservative tests with good properties for small sample sizes, we propose a constrained bootstrap test in Sect. 3. Section 4 is devoted to a small simulation study illustrating good finite sample properties of the bootstrap confidence intervals and tests. Finally, all proofs and technical details are given in appendix.

2 Comparing the area between the curves

We consider two independent samples of n_1 and n_2 observations. In each group ($\ell = 1, 2$), there exist k_ℓ different covariates, say $x_{\ell,1}, \dots, x_{\ell,k_\ell}$, such that at each covariate $x_{\ell,i}$ $n_{\ell,i}$ independent identically distributed (iid) observations $\{Y_{\ell,i,j} : j = 1, \dots, n_{\ell,i}\}$ are available such that $n_\ell = \sum_{i=1}^{k_\ell} n_{\ell,i}$ ($\ell = 1, 2$). The total sample size is denoted by $n = n_1 + n_2$. We assume that the covariates vary in a set $\mathcal{X} \subset \mathbb{R}^d$ (for some $d \in \mathbb{N}$) and that the relation between the covariates and responses can be represented by a nonlinear regression model of the form

$$Y_{\ell,i,j} = m_\ell(x_{\ell,i}, \beta_\ell) + \eta_{\ell,i,j}, \quad j = 1, \dots, n_{\ell,i}, \quad i = 1, \dots, k_\ell, \tag{2}$$

where $m_\ell(\cdot, \beta_\ell) \in \ell^\infty(\mathcal{X})$ is the regression function with parameters $\beta_\ell \in \mathbb{R}^{p_\ell}$, $p_\ell \in \mathbb{N}$ and $\ell^\infty(\mathcal{X})$ denotes the space of bounded real-valued functions $f : \mathcal{X} \to \mathbb{R}$.

In model (2), the errors $\{\eta_{\ell,i,j} : j = 1, \dots, n_{\ell,i},\ i = 1, \dots, k_\ell\}$ denote iid random variables with mean 0 and variance $\sigma_\ell^2 > 0$, $\ell = 1, 2$.

In this paper, we are interested in the similarity between the regression functions m_1 and m_2, where the distance between the two functions is measured by the L^1-distance. More precisely, we consider the L^1-distance

$$d_1 = d_1(\beta_1, \beta_2) := \int_{\mathcal{X}} |m_1(x, \beta_1) - m_2(x, \beta_2)| \mathrm{d}x, \tag{3}$$

and develop confidence intervals for d_1 and statistical tests for the hypotheses

$$H_0 : d_1(\beta_1, \beta_2) \geq \epsilon \quad \text{vs.} \quad H_1 : d_1(\beta_1, \beta_2) < \epsilon, \tag{4}$$

where $\epsilon > 0$ is a pre-specified constant. Note that the rejection of H_0 in (4) allows to decide at a controlled type I error that the area between the two curves is smaller than a given threshold.

As pointed out in Introduction, the choice of L^1-distance for the comparison of curves poses several mathematical challenges, which are caused by the fact that in contrast to the L^2-norm, the mapping $f \to \int_{\mathcal{X}} |f(x)| \mathrm{d}x$ from $l^\infty(\mathcal{X})$ onto $\mathbb{R}$ is in general not (Hadamard-)differentiable. Our approach for investigating the similarity between the regression functions m_1 and m_2 is based on a direct estimate of the distance $d_1(\beta_1, \beta_2)$. To be precise, let $\hat{\beta}_1$ and $\hat{\beta}_2$ denote appropriate estimates of the parameters in model $m_1(\cdot, \beta_1)$ and $m_2(\cdot, \beta_2)$ obtained from the samples $\{Y_{1,i,j} | j = 1, \dots, n_{1,i},\ i = 1, \dots, k_1, \}$ and $\{Y_{2,i,j} | j = 1, \dots, n_{2,i},\ i = 1, \dots, k_2, \}$, respectively. (Precise assumptions on the properties of these estimates are given in appendix.) The estimate of the area between the curves d_1 in (3) is then defined by

$$\hat{d}_1 = d_1(\hat{\beta}_1, \hat{\beta}_2) = \int_{\mathcal{X}} |m_1(x, \hat{\beta}_1) - m_2(x, \hat{\beta}_2)| \mathrm{d}x, \tag{5}$$

and we will show that the normalized statistic $\sqrt{n}(\hat{d}_1 - d_1)$ converges weakly to a random variable T with non-degenerate distribution. For this purpose, we introduce the set

$$\mathcal{N} := \{x \in \mathcal{X} |\ m_1(x, \beta_1) - m_2(x, \beta_2) = 0\} \tag{6}$$

as the set of points, where the two regression functions coincide. Throughout this paper, the symbol $\xrightarrow{d}$ means weak convergence in distribution.

Theorem 1 *Suppose that Assumptions* 1–7 *in appendix are satisfied, in particular* $n_1, n_2 \to \infty$ *such that* $n/n_1 \to \kappa \in (1, \infty)$ *and* $n_{\ell,i}/n_\ell \to \zeta_{\ell,i} \in (0, 1)$ *for* $i = 1, \dots, k_\ell$, $\ell = 1, 2$. *The statistic* $\hat{d}_1$ *defined in* (5) *satisfies*

$$\begin{aligned} \sqrt{n}(\hat{d}_1 - d_1) \xrightarrow{d} T := \int_{\mathcal{N}^c} & \operatorname{sgn}(m_1(x, \beta_1) \\ & - m_2(x, \beta_2))\, G(x)\, \mathrm{d}x + \int_{\mathcal{N}} |G(x)|\, \mathrm{d}x \end{aligned} \tag{7}$$

where $\{G(x)\}_{x\in\mathcal{X}}$ is a centred Gaussian process in $\ell^\infty(\mathcal{X})$ defined by

$$G(x) = \Big(\frac{\partial}{\partial b_1} m_1(x,b_1)\Big|_{b_1=\beta_1}\Big)^\top \sqrt{\kappa}\,\Sigma_1^{-1/2} Z_1 - \Big(\frac{\partial}{\partial b_2} m_2(x,b_2)\Big|_{b_2=\beta_2}\Big)^\top \sqrt{\frac{\kappa}{\kappa-1}}\,\Sigma_2^{-1/2} Z_2,$$

Z_1 and Z_2 are p_1 and p_2-dimensional standard normal distributed random variables, respectively, and the matrices Σ_1 and Σ_2 are defined by

$$\Sigma_\ell = \frac{1}{\sigma_\ell^2}\sum_{i=1}^{k_\ell} \zeta_{\ell,i}\Big(\frac{\partial}{\partial b_\ell} m_\ell(x_{\ell,i},b_\ell)\Big|_{b_\ell=\beta_\ell}\Big)\Big(\frac{\partial}{\partial b_\ell} m_\ell(x_{\ell,i},b_\ell)\Big|_{b_\ell=\beta_\ell}\Big)^\top \quad (l=1,2).$$

If the distribution of T in (7) would be known and q_α denotes the corresponding α -quantile, it follows from Theorem 1 that the coverage probability of the "oracle" confidence interval $[0, \hat{d}_1 - \frac{q_\alpha}{\sqrt{n}})$ for the L^1-distance d_1 converges with increasing sample size to $1-\alpha$. Similarly, a simple calculation shows that the test, which rejects the null hypothesis, whenever $\hat{d}_1 < \epsilon + \frac{q_\alpha}{\sqrt{n}}$, is a consistent asymptotic level α test for the hypotheses (4).

However, the distribution of the limiting random variable T in (7) is not easily accessible, because it depends on certain unknown nuisance parameters, in particular on the set $\mathcal{N}$ of points, where the two functions m_1 and m_2 coincide. The unknown covariance matrices Σ_1 and Σ_2, which essentially define the Gaussian process $\{G(x)\}_{x\in\mathcal{X}}$, can be estimated in a rather straightforward manner by

$$\hat{\Sigma}_\ell = \frac{1}{\hat{\sigma}_\ell^2}\sum_{i=1}^{k_\ell} \frac{n_{\ell,i}}{n_\ell}\Big(\frac{\partial}{\partial b_\ell} m_\ell(x_{\ell,i},b_\ell)\Big|_{b_\ell=\hat{\beta}_\ell}\Big)\Big(\frac{\partial}{\partial b_\ell} m_\ell(x_{\ell,i},b_\ell)\Big|_{b_\ell=\hat{\beta}_\ell}\Big)^\top, \tag{8}$$

where $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are consistent estimators of the variances σ_1^2 and σ_2^2, respectively. On the other hand, the estimation of the set $\mathcal{N}$ is more difficult. For a constant $c>0$, we define an estimate by

$$\hat{\mathcal{N}} = \Big\{x\in\mathcal{X} \;\Big|\; |m_1(x,\hat{\beta}_1) - m_2(x,\hat{\beta}_2)| < c\sqrt{\frac{\log n}{n}}\Big\}. \tag{9}$$

Consequently, let $\hat{G}$ denote the process G introduced in Theorem 1, where the parameter β_ℓ and the matrix Σ_ℓ have been replaced by their estimates $\hat{\beta}_l$ and $\hat{\Sigma}_\ell$, respectively, and κ by n/n_1. Then we define the random variable

$$\hat{T} := \int_{\hat{\mathcal{N}}^c} \operatorname{sgn}(m_1(x,\hat{\beta}_1) - m_2(x,\hat{\beta}_2))\,\hat{G}(x)\,\mathrm{d}x + \int_{\hat{\mathcal{N}}} |\hat{G}(x)|\,\mathrm{d}x \tag{10}$$

and denote by $\hat{q}_{0,\alpha}$ the α-quantile of the corresponding distribution conditional on $\hat{\beta}_1, \hat{\beta}_2$, which can easily be simulated. We now define a confidence interval

$$\hat{I}_n := \Big[0, \hat{d}_1 - \frac{\hat{q}_{0,\alpha}}{\sqrt{n}}\Big), \tag{11}$$

for the L^1-distance d_1, and using the duality between confidence intervals and statistical tests (see, for example, Aitchison 1964), we propose to reject the null hypothesis in (4), whenever $\hat{I}_n \subset [0, \epsilon)$, which is equivalent to

$$\hat{d}_1 < \epsilon + \frac{\hat{q}_{0,\alpha}}{\sqrt{n}}\ . \tag{12}$$

We then obtain the following theorem on the significance level and the power of the above procedure.

Theorem 2 *If the assumptions of Theorem 1 are satisfied, then* (11) *defines an asymptotic confidence interval for the quantity* d_1, *i.e.*

$$\lim_{n\to\infty} \mathbb{P}\big(d_1 \in \hat{I}_n\big) = \lim_{n\to\infty} \mathbb{P}\Big(d_1 \in \Big[0, \hat{d}_1 - \frac{\hat{q}_{0,\alpha}}{\sqrt{n}}\Big)\Big) = 1 - \alpha.$$

Moreover, the test defined in (12) *is a consistent and asymptotic level* α*-test, i.e.*

(i) *If* $d_1 \geq \epsilon$, *then* $\limsup_{n\to\infty} \mathbb{P}\big(\hat{d}_1 < \epsilon + \frac{\hat{q}_{0,\alpha}}{\sqrt{n}}\big) \leq \alpha$

(ii) *If* $d_1 < \epsilon$, *then* $\liminf_{n\to\infty} \mathbb{P}\big(\hat{d}_1 < \epsilon + \frac{\hat{q}_{0,\alpha}}{\sqrt{n}}\big) = 1.$

Remark 1 A two-sided (asymptotic) confidence interval for d_1 is given by $[\hat{d}_1 - \frac{\hat{q}_{0,1-\alpha/2}}{\sqrt{n}},\ \hat{d}_1 - \frac{\hat{q}_{0,\alpha/2}}{\sqrt{n}}]$.

3 Bootstrap methodology

The confidence interval and test proposed in Sect. 2 require the precise estimation of the set $\mathcal{N}$ in (6), which might be difficult for small sample sizes. In order to address this problem and to obtain also a better approximation of the nominal level, we propose a parametric bootstrap approach. While this is quite standard for the construction of confidence intervals we will develop and investigate a novel constrained bootstrap approach for testing the hypotheses (4). This approach will result in more powerful tests compared to the confidence interval approach (see Sect. 4). We describe the construction of a bootstrap confidence interval in Algorithm 1.

Algorithm 1 Bootstrap confidence interval

(1) Calculate the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ (for each group).

(2) For $\ell = 1, 2, i = 1, \dots, k_\ell, j = 1, \dots, n_{\ell,i}$ generate bootstrap data from the model

$$Y^*_{\ell,i,j} = m_\ell(x_{\ell,i}, \hat{\beta}_\ell) + \eta^*_{\ell,i,j}, \tag{13}$$

where the errors $\eta^*_{\ell,i,j}$ are independent centered normal distributed with variance $\hat{\sigma}^2_\ell$.

(3) Let $\hat{\beta}^*_1$ and $\hat{\beta}^*_2$ denote the estimators of β_1 and β_2 from the bootstrap data in (13). Calculate the bootstrap test statistic $\hat{d}^*_1 := d_1(\hat{\beta}^*_1, \hat{\beta}^*_2)$ and define $\hat{q}^*_{1-\alpha,0}$ as the $(1-\alpha)$-quantile of the distribution of $\hat{d}^*_1$.

(4) The bootstrap confidence interval is defined by

$$\hat{I}^*_n := \big[0, \hat{q}^*_{1-\alpha,0}\big). \tag{14}$$

As in Sect. 2, the duality between confidence intervals and statistical tests (see Aitchison 1964) yields a test for the hypotheses (4), which rejects the null hypothesis, whenever $\hat{I}^*_n \subset [0, \epsilon)$, that is,

$$\hat{q}^*_{1-\alpha,0} < \epsilon \tag{15}$$

The following result shows that this ad hoc approach yields a consistent asymptotic level α test.

Theorem 3 *In addition to the assumptions of Theorem 1, assume that $\lambda(\mathcal{N}) = 0$. The interval $\hat{I}^*_n$ in (14) defines an asymptotic confidence interval for the quantity d_1, i.e.*

$$\lim_{n\to\infty} \mathbb{P}\big(d_1 \in \hat{I}^*_n\big) = 1 - \alpha.$$

Moreover, the test defined in (15) is a consistent and asymptotic level α-test, i.e.

(i) *If $d_1 \geq \epsilon$ then* $\limsup_{n\to\infty} \mathbb{P}\big(\hat{q}^*_{1-\alpha,0} < \epsilon\big) \leq \alpha$

(ii) *If $d_1 < \epsilon$ then* $\liminf_{n\to\infty} \mathbb{P}\big(\hat{q}^*_{1-\alpha,0} < \epsilon\big) = 1.$

The finite sample properties of the confidence interval (14) and the test (15) will be investigated in Sect. 4. In particular, it will be demonstrated that, by its construction, the test (15) is rather conservative and not very powerful. Therefore, we propose as an alternative a constrained bootstrap test for the hypotheses (4), which addresses the specific structure of the composite hypotheses (4). The pseudocode for this test is summarized in Algorithm 2.

Algorithm 2 Constrained parametric bootstrap test

(1) Calculate the test statistic $\hat{d}_1 := d_1(\hat{\beta}_1, \hat{\beta}_2)$ defined in (5).

(2) Calculate the estimators $\tilde{\beta}_1$ and $\tilde{\beta}_2$ of the parameters β_1 and β_2, respectively, under the additional constraint that $d_1(\beta_1, \beta_2) = \epsilon$. Define

$$\hat{\hat{\beta}}_\ell = \begin{cases} \hat{\beta}_\ell, & \hat{d}_1 \geq \epsilon \\ \tilde{\beta}_\ell, & \hat{d}_1 < \epsilon. \end{cases}, \quad \ell = 1, 2. \tag{16}$$

(3) For $\ell = 1, 2, i = 1, \ldots, k_\ell, j = 1, \ldots, n_{\ell,i}$ generate bootstrap data from the model

$$Y^*_{\ell,i,j} = m_\ell(x_{\ell,i}, \hat{\hat{\beta}}_\ell) + \eta^*_{\ell,i,j}, \tag{17}$$

where the errors $\eta^*_{\ell,i,j}$ are independent centered normal distributed with variance $\hat{\sigma}^2_\ell$.

(4) Let $\hat{\beta}^*_1$ and $\hat{\beta}^*_2$ denote the estimators of β_1 and β_2 from the bootstrap data in (17). Calculate the bootstrap test statistic $\hat{d}^*_1 := d_1(\hat{\beta}^*_1, \hat{\beta}^*_2)$ and define $\hat{q}^*_{\alpha,1}$ as the α-quantile of the distribution of $\hat{d}^*_1$.

(5) The null hypothesis in (4) is rejected, whenever

$$\hat{d}_1 < \hat{q}^*_{\alpha,1}. \tag{18}$$

Remark 2 (i) Note that, by definition (16), $\hat{\hat{d}}_1 := d_1(\hat{\hat{\beta}}_1 \hat{\hat{\beta}}_2) \geq \epsilon$. Therefore, Algorithm 2 generates in step (3) bootstrap data under the null hypothesis.

(ii) In practice, the quantile can be estimated with arbitrary precision generating bootstrap replicates $\hat{d}^{*(1)}_1, \ldots, \hat{d}^{*(B)}_1$ as described in steps (3) and (4) in Algorithm 2 and calculating the empirical α-quantile, say $\hat{q}^{(B)}_{\alpha,1}$, from this sample.

(iii) The following result shows that this parametric bootstrap test has asymptotic level α and is consistent if the set $\mathcal{N}$ defined in (6) has Lebesgue measure 0.

Theorem 4 *In addition to Assumption* 1–7 *in appendix, assume that the set* $\mathcal{N}$ *defined in* (6) *has Lebesgue measure zero. Further assume that the α-quantile* q_α *of the random variable T in* (7) *is negative. The constrained bootstrap test defined by* (17) *in Algorithm* 2 *has asymptotic level α and is consistent. More precisely*,

(i) *If the null hypothesis* $H_0 : d_1 \geq \epsilon$ *holds, then* $\lim_{n_1,n_2\to\infty} \mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1}) = 0$ *if* $d_1 > \epsilon$ *and* $\lim_{n_1,n_2\to\infty} \mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1}) = \alpha$, *if* $d_1 = \epsilon$.

(ii) *If the alternative hypothesis* $H_1 : d_1 < \epsilon$ *holds, then* $\lim_{n_1,n_2\to\infty} \mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1}) = 1$.

Remark 3 (i) The assumption that q_α be negative is needed to show that $\mathbb{P}(\hat{\hat{d}} < \hat{q}^*_{\alpha,1}) \to 0$ in the interior of H_0, that is, $d_1 > \epsilon$. If $\mathcal{N}$ has Lebesgue measure 0 the distribution of T is Gaussian and it reduces to $\alpha < 0.5$.

(ii) Algorithms 1 and 2 have been formulated under the assumption of homoscedastic errors. However, these algorithms can easily be modified to address heteroscedasticity, for example, by using a different estimate of the asymptotic variance.

Remark 4 (i) Let $\theta = \{m_1(x,\beta_1) - m_2(x,\beta_2)\}_{x\in\mathcal{X}}$, $\hat{\theta} = \{m_1(x,\hat{\hat{\beta}}_1) - m_2(x,\hat{\hat{\beta}}_2)\}_{x\in\mathcal{X}}$ and $\hat{\theta}^* = \{m_1(x,\hat{\beta}_1^*) - m_2(x,\hat{\beta}_2^*)\}_{x\in\mathcal{X}}$ denote processes on $\ell^\infty(\mathcal{X})$ and define the mapping $\Phi(f) = \int_{\mathcal{X}} |f(x)| \mathrm{d}x$ from $\ell^\infty(\mathcal{X})$ onto $\mathbb{R}$. By the proof of Theorem 1, we have

$$\sup_{h\in BL} \left|\mathbb{E}\left[h\left(\sqrt{n}(\Phi(\hat{\theta}) - \Phi(\theta))\right)\right] - \mathbb{E}\left[h\left(\Phi'_\theta(\mathbb{G})\right)\right]\right| = o(1)\,,$$

where BL denotes the space of bounded Lipschitz functions (see Van der Vaart 2000), $\mathbb{G} = \{G(x)\}_{x\in\mathcal{X}}$ is the Gaussian process defined in Theorem 1 and Φ'_θ denotes the directional Hadamard derivative at the process $\theta \in \ell^\infty(\mathcal{X})$. Now, according to Theorem 3.1 in Fang and Santos (2019), the corresponding statement for the bootstrap process

$$\sup_{h\in BL} \left|\mathbb{E}^*\left[h\left(\sqrt{n}(\Phi(\hat{\theta}^*) - \Phi(\hat{\theta}))\right)\right] - \mathbb{E}\left[h\left(\Phi'_\theta(\mathbb{G})\right)\right]\right| = o_{\mathbb{P}}(1)$$

(here $\mathbb{E}^*$ denote the expectation conditional on the sample) holds if and only if the directional derivative of the mapping Φ at θ is linear, that is, Φ is Hadamard differentiable at θ. However, by the proof of Theorem 1, this is the case if and only if $\lambda(\mathcal{N}) = 0$.

From a practical point of view, the condition $\lambda(\mathcal{N}) = 0$ will be fulfilled in most applications. For example, if the predictor is one-dimensional the curves corresponding to typically used parametric regression models m_1 and m_2 either intersect in at most one point or they are completely identical (which is unlikely in most applications). Therefore, Theorem 4 is applicable in most applications and ensures that the parametric bootstrap defined by Algorithm 2 yields a statistically valid procedure.

(ii) A bootstrap procedure for which consistency can be proved even in the case $\lambda(\mathcal{N}\Rightarrow > 0$ can be obtained using the results of Fang and Santos (2019); in particular, we will construct an estimator of the directional derivative of the L_1-norm mapping that fulfils the assumptions of Theorem 3.2 in their paper. To be precise, we define

$$\hat{\Phi}'(f) := \int_{|\hat{\theta}|\geq 1/s_n} \mathrm{sgn}(\hat{\theta}(x)) f(x)\,\mathrm{d}x + \int_{|\hat{\theta}|<1/s_n} |f(x)|\,\mathrm{d}x$$

for some sequence s_n satisfying $s_n/\sqrt{n} \to 0$. We then proceed as in Algorithm 2, where the steps (4) and (5) are replaced by

(4’) Calculate the bootstrap test statistic $\hat{\Phi}'(\hat{\theta}^* - \hat{\theta})$ and the α-quantile $\hat{q}^*_{\alpha,1}$ of its distribution.

(5') The null hypothesis in (4) is rejected, whenever $\hat{d}_1 < \hat{q}^*_{\alpha,1} + \epsilon$. It can then be shown (see Appendix) that the statements of Theorem 4 hold for this test, even in the case $\lambda(\mathcal{N}) > 0$.

4 Finite sample properties

We investigate the finite sample properties of the confidence intervals and the tests for the hypotheses (4) by means of a small simulation study. For this purpose, we consider two E-max models

$$m_\ell(x, \beta_\ell) = \beta_{\ell 1} + \frac{\beta_{\ell 2} x}{\beta_{\ell 3} + x} \quad (\ell = 1, 2), \tag{19}$$

where $x \in \mathcal{X} = [0, 4]$. We consider two scenarios for the parameters:

$$\text{Intersecting curves}: \quad \beta_1 = (5, 3, 1)^\top, \ \ \beta_2 = (5, 3+\gamma, 1+\gamma)^\top, \quad \gamma \geq 0 \tag{20}$$

$$\text{Parallel curves}: \quad \beta_1 = (\delta, 5, 1)^\top, \ \ \beta_2 = (0, 5, 1)^\top, \quad \delta \geq 0. \tag{21}$$

Some typical curves are displayed in Fig. 1.

An equal number of observations is allocated to five equidistant dose levels $x_{1,1} = x_{2,1} = 0, x_{1,2} = x_{2,2} = 1, \dots, x_{1,5} = x_{2,5} = 4$, and the sample sizes for both groups are given by $n_1 = n_2 = 20, 50, 100$ and 200. The errors in the regression models (2) are centred normal distributions with variances chosen as $(\sigma_1^2, \sigma_2^2) = (0.25, 0.25)$ and $(0.25,\ 0.5)$. All bootstrap results are obtained by $B = 300$ replications.

As estimators for the parameters in the regression models (2), we use least squares estimators, that is,

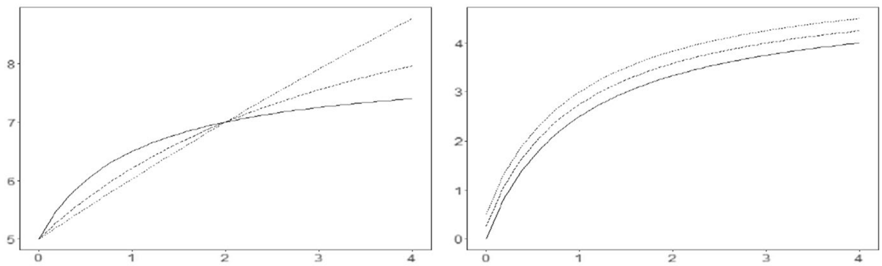

Fig. 1 Typical E-max curves considered in the simulation study. Left panel: intersecting curves defined by (20) with $\gamma = 0$ (solid), $\gamma = 2.7$ (dashed) and $\gamma = 30$ (dotted). Right panel: parallel curves defined by (21) with $\delta = 0$ (solid), $\delta = 1/4$ (dashed) and $\delta = 1/2$ (dotted)

$$
\begin{aligned}
\hat{\beta}_\ell =& \arg\min_{\beta_\ell \in \mathbb{R}^{p_\ell}} \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} (Y_{\ell,i,j} - m_\ell(x_{\ell,i}, \beta_\ell))^2 , \\
\hat{\sigma}_\ell^2 =& \frac{1}{n_\ell} \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} (Y_{\ell,i,j} - m_\ell(x_{\ell,i}, \hat{\beta}_\ell))^2
\end{aligned}
$$

($\ell = 1, 2$). We start investigating the coverage probabilities of the asymptotic and bootstrap confidence intervals for the distance d_1 defined in (11) and (14), respectively. For the asymptotic confidence interval, we estimate the set $\mathcal{N}$ by (9) with $c = 1$. The parameters in the two E-max models (19) are defined by (20) and (21) such that $d_1 = 1$. The corresponding results are given in Table 1, where the upper part corresponds to the intersecting and the lower part to the parallel scenario. We observe that the coverage probabilities of the asymptotic confidence interval are too small, but they improve with increasing sample size. The results for the bootstrap confidence intervals are more satisfactory. For small sample sizes, the bootstrap yields intervals with a too large coverage probability, but in general it provides an improvement.

Next, we consider the problem of testing the hypotheses (4) with $\epsilon = 1$. We begin with the two tests (12) and (15), which have been derived from the asymptotic and bootstrap confidence interval, respectively. The corresponding rejection probabilities are displayed in Fig. 2 for the E-max models (19) with parameters (21) (parallel curves). Note that in this case $d_1 = 4\delta$ and $\lambda(\mathcal{N}) = 0$, whenever $\delta \neq 0$. Therefore, the cases $\delta \geq 0.25$ and $\delta < 0.25$ correspond to the null hypothesis and alternative in (4), respectively. The horizontal solid line marks the significance level $\alpha = 0.05$, and the vertical solid line corresponds to the boundary of the null hypothesis, that is, $d_1 = 1$. The curves reflect the qualitative behaviour predicted by Theorems 2 and 3. Note that the power curves are decreasing in d_1 as the null hypothesis is given by $H_0 : d_1 \geq 1$. We observe that the asymptotic test (12) does not keep its nominal level. In particular, for small sample sizes the level

Table 1 Coverage probabilities of the asymptotic (left) and bootstrap 95%-confidence interval (right)

(n_1, n_2)	(σ_1^2, σ_2^2)			
	(0.25, 0.25)	(0.25, 0.5)	(0.25, 0.25)	(0.25, 0.5)
(20,20)	0.640	0.680	1.000	1.000
(50,50)	0.690	0.695	0.990	1.000
(100,100)	0.830	0.810	0.965	0.990
(200,200)	0.845	0.845	0.955	0.960
(20,20)	0.620	0.645	1.000	1.000
(50,50)	0.695	0.760	0.994	1.000
(100,100)	0.775	0.755	0.972	0.984
(200,200)	0.855	0.840	0.938	0.960

The regression functions are given by (19) such that $d_1 = 1$. Upper part: intersecting curves defined by (20). Lower part: parallel curves defined by (21)

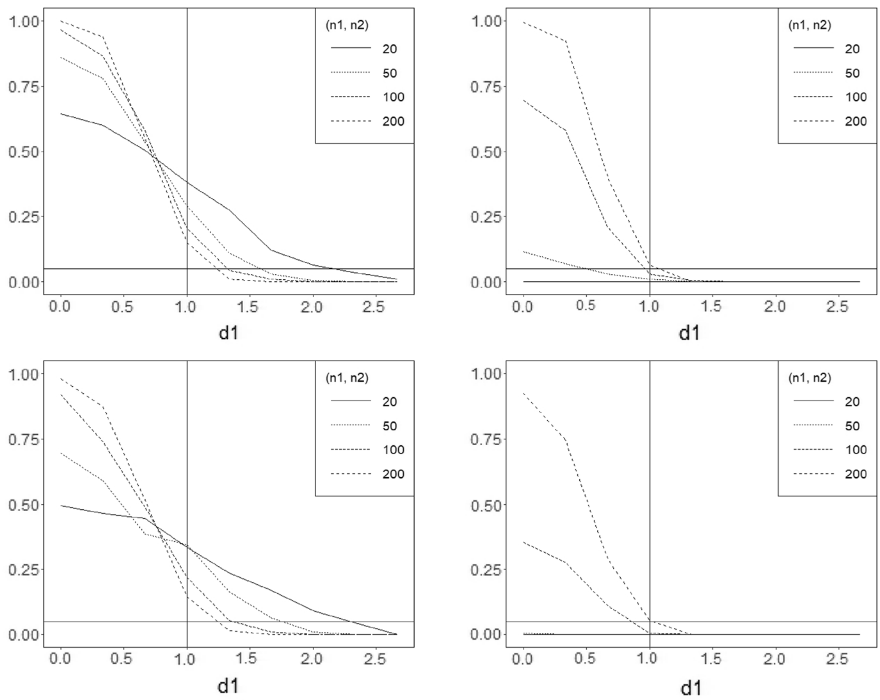

Fig. 2 Rejection probabilities of the tests (12) (left) and (15) (right) for the hypotheses (4). First row: $\sigma_1^2 = \sigma_2^2 = 0.25$; Second row: $\sigma_1^2 = 0.25$, $\sigma_2^2 = 0.5$. The nominal level is $\alpha = 0.05$, and parallel E-Max models defined by (21) are considered

is substantially exceeded. On the other hand, the bootstrap test (15) keeps the nominal level in all situations under consideration. The asymptotic test has more power, but this advantage comes at the cost of an unreliable approximation of the nominal level. Therefore, if one would have to choose from these tests, we would recommend to use the test based on the bootstrap confidence interval. Note that this test has not much power for the sample sizes $n_1 = n_2 = 20$ and 50, but the constrained bootstrap test developed in Sect. 3 will yield a further improvement.

In Fig. 3, we illustrate the performance of the test (18) (constrained bootstrap - see Algorithm 2) for testing the hypotheses (4) with $\epsilon = 1$. The rejection probabilities for the situation investigated in Fig. 2 (parallel E-Max models defined by (21)) are shown in the right part of the figure. These results are directly comparable with the right panels in Fig. 2. We observe that the constrained bootstrap test (18) yields a substantial improvement in power compared to the test (15), which is based on the bootstrap confidence interval. For example, if $d_1 = 0$, $n_1 = n_2 = 50$, $\sigma_1^2 = \sigma_2^2 = (0.25, 0.25)$, the test (15) has approximately power 0.115, while the power of the constrained bootstrap test is 0.455. The left part of the Fig. 3 shows the results for intersecting E-Max models (defined by (20)). A comparison with

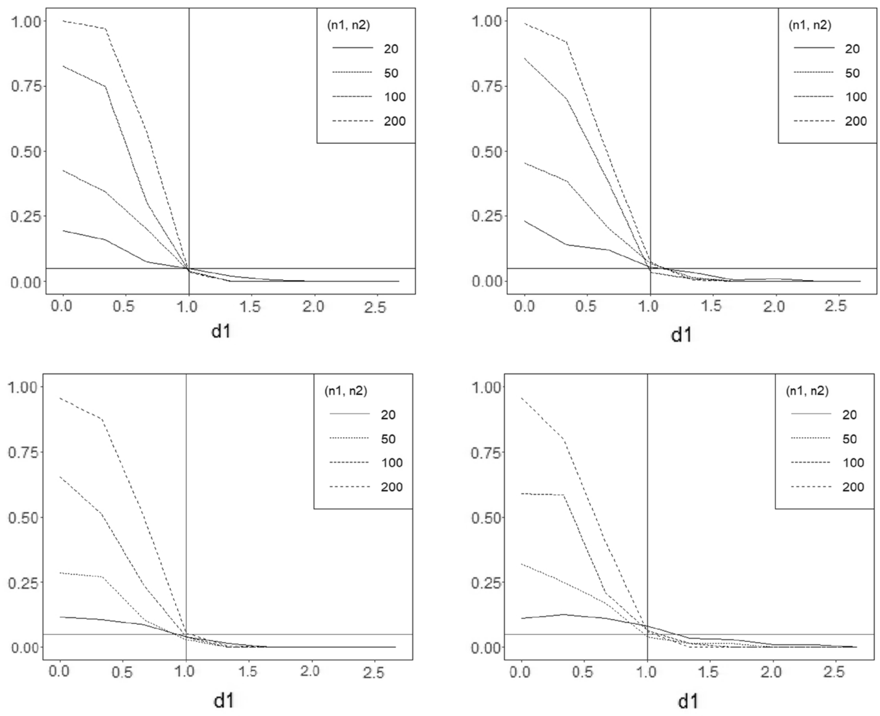

Fig. 3 Rejection probabilities of the constrained bootstrap test (18) (Algorithm 2) for the hypotheses (4). Left panels: intersecting E-max models with parameters (20) Right panels: parallel E-max models with parameters (21). First row: $\sigma_1^2 = \sigma_2^2 = 0.25$; Second row: $\sigma_1^2 = 0.25, \sigma_2^2 = 0.5$. The nominal level is $\alpha = 0.05$

the right part shows that the differences in power between the two cases (intersecting E-Max and shifted E-Max curves) are rather small.

Acknowledgements The authors would like to thank two anonymous referees for their constructive comments on an earlier version of this paper. This research is supported by the European Union through the European Joint Programme on Rare Diseases under the European Union's Horizon 2020 Research and Innovation Programme Grant Agreement Number 825575.

Appendix

In this appendix, we give proofs to all theoretical results in this paper. For this purpose, we require the following assumptions:

Assumption 1 The errors $\eta_{\ell,i,j}$ have finite variance σ_ℓ^2 and mean 0.

Assumption 2 The covariate region $\mathcal{X} \subset \mathbb{R}^d$ is compact and the number and location levels of k_ℓ does not depend on n_ℓ for $\ell = 1, 2$.

Assumption 3 All estimators of the parameters β_1, β_2 are computed over compact sets $B_1 \subset \mathbb{R}^{p_1}$ and $B_2 \subset \mathbb{R}^{p_2}$.

Assumption 4 The regression functions m_1 and m_2 are twice continuously differentiable with respect to the parameters for all b_1, b_2 in the neighbourhoods of the true parameters β_1, β_2 and all $x \in \mathcal{X}$. The functions $(x, b_\ell) \rightarrowtail m_\ell(x, b_\ell)$ and their first two derivatives are continuous on $\mathcal{X} \times B_\ell$ for $\ell = 1, 2$.

Assumption 5 Defining

$$\psi^{(n)}_{a,\ell}(b) := \sum_{i=1}^{k_\ell} \frac{n_{\ell,i}}{n_\ell}(m_\ell(x_{\ell,i}, a) - m_\ell(x_{\ell,i}, b))^2,$$

we assume that for any $u > 0$ there exists a constant $v_{u,\ell} > 0$ such that

$$\liminf_{n\to\infty} \inf_{a\in B_\ell} \inf_{|b-a|\geq u} \psi^{(n)}_{a,\ell}(b) \geq v_{u,\ell}, \quad \ell = 1, 2.$$

Assumption 6 The matrices Σ_ℓ are non-singular and the sample sizes n_1, n_2 converge to infinity such that

$$\lim_{n_\ell\to\infty} \frac{n_{\ell,i}}{n_\ell} = \xi_{\ell,i} > 0, \quad i = 1, \dots, k_\ell, \ \ell = 1, 2,$$

and

$$\lim_{n_1,n_2\to\infty} \frac{n}{n_1} = \kappa \in (1, \infty).$$

Assumption 7 We denote by $\hat{\beta}_1, \hat{\beta}_2$ estimators of the parameters β_1, β_2 and assume that they can be linearized, meaning the estimators fulfil the following condition:

$$\sqrt{n_\ell}(\hat{\beta}_\ell - \beta_\ell) = \frac{1}{\sqrt{n_\ell}} \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} \phi_{\ell,i,j} + o_{\mathbb{P}}(1) \ \text{ as } n_\ell \to \infty, \ \ell = 1, 2$$

with square integrable influence functions $\phi_{1,i,j}$ and $\phi_{2,i,j}$ satisfying

$$\mathbb{E}[\phi_{\ell,i,j}] = 0, \ j = 1, \dots, n_{\ell,i}, \ i = 1, \dots, k_\ell, \ \ell = 1, 2.$$

This implies that the asymptotic distribution of $\hat{\beta}_1$ and $\hat{\beta}_2$ is given by

$$\sqrt{n_\ell}(\hat{\beta}_\ell - \beta_\ell) \xrightarrow{d} \mathcal{N}(0, \Sigma_\ell^{-1}), \ \ell = 1, 2,$$

where the asymptotic covariance matrix is given by

$$\Sigma_\ell^{-1} = \sum_{i=1}^{k_\ell} \xi_{\ell,i} \mathbb{E}[\phi_{\ell,i,j}\phi_{\ell,i,j}^\top],\ \ell = 1, 2.$$

Moreover, the variance estimators $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ used in (8) are consistent.

A.1 Proof of Theorem 1

We will prove this result by an application of the (functional) delta method for *directionally* differentiable functionals as stated in Theorem 2.1 in Shapiro (1991). We introduce the notations $\theta(x) = m_1(x, \beta_1) - m_2(x, \beta_2)$, $\hat{\theta}(x) = m_1(x, \hat{\beta}_1) - m_2(x, \hat{\beta}_2)$, $\theta = \{\theta(x)\}_{x\in\mathcal{X}}$ and $\hat{\theta} = \{\hat{\theta}(x)\}_{x\in\mathcal{X}}$, and will show below that the mapping

$$\Phi : \begin{cases} \ell^\infty(\mathcal{X}) \to \mathbb{R} \\ f \to \Phi(f) = \int_\mathcal{X} |f(x)| \mathrm{d}x \end{cases}$$

is *directionally* Hadamard differentiable with respect to $(\ell^\infty(\mathcal{X}), \|\cdot\|_1)$ and the absolute value norm on $\mathbb{R}$, where the derivative is given by

$$\Phi_h' : \begin{cases} \ell^\infty(\mathcal{X}) \to \mathbb{R} \\ f \to \Phi_h'(f) = \int_{\{h\neq 0\}} \mathrm{sgn}(h(x)) f(x) \mathrm{d}x + \int_{\{h=0\}} |f(x)| \mathrm{d}x \end{cases}$$

at $h \in \ell^\infty(\mathcal{X})$. Note that $\big(\ell^\infty(\mathcal{X}), \|\cdot\|_1\big)$ is still separable and that its norm is weaker than the sup-norm.

Hence, the convergence in distribution

$$\sqrt{n}\big\{\hat{\theta}(x) - \theta(x)\big\}_{x\in\mathcal{X}} \xrightarrow{d} \{G(x)\}_{x\in\mathcal{X}}$$

in $\big(\ell^\infty(\mathcal{X}), \|\cdot\|_\infty\big)$ established in Dette et al. (2018) is also valid in this setting. In particular, applying the (directional) delta method (Theorem 2.1 in Shapiro 1991) gives

$$\begin{aligned} \sqrt{n}(\hat{d}_1 - d_1) &= \sqrt{n} \int_\mathcal{X} |\hat{\theta}(x)| - |\theta(x)| \, \mathrm{d}x \\ &= \sqrt{n}\Big(\Phi(\{\hat{\theta}(x)\}_{x\in\mathcal{X}}) - \Phi(\{\theta(x)\}_{x\in\mathcal{X}})\Big) \\ &\xrightarrow{d} \Phi_\theta'(\{G(x)\}_{x\in\mathcal{X}}) \\ &= \int_{\mathcal{N}^c} \mathrm{sgn}(\theta(x))\, G(x)\, \mathrm{d}x + \int_\mathcal{N} |G(x)| \, \mathrm{d}x, \end{aligned}$$

where $\mathcal{N}$ is defined in (6). Therefore, we are left with showing the differentiability of the functional Φ. For this purpose, we write $\Phi = \Phi_1 \circ \Phi_2$, where

$$\Phi_1 : \begin{cases} \ell^\infty(\mathcal{X}) \to \mathbb{R} \\ f \to \Phi_1(f) = \int_\mathcal{X} f(x) \mathrm{d}x \end{cases}, \quad \Phi_2 : \begin{cases} \ell^\infty(\mathcal{X}) \to \ell^\infty(\mathcal{X}) \\ \Phi_2(f) = |f|. \end{cases}$$

As a linear mapping Φ_1 is obviously Hadamard differentiable with derivative $(\Phi_1)'_h = \Phi_1$ at $h \in \ell^\infty(\mathcal{X})$, with respect to $(\ell^\infty(\mathcal{X}), \|\cdot\|_1)$. We prove below that Φ_2 is directionally Hadamard differentiable with respect to $(\ell^\infty(\mathcal{X}), \|\cdot\|_1)$ and derivative

$$(\Phi_2)'_h : \begin{cases} \ell^\infty(\mathcal{X}) \to \ell^\infty(\mathcal{X}) \\ f \to (\Phi_2)'_h(f) = \mathbb{1}_{\{h\neq 0\}}\mathrm{sgn}(h)f + \mathbb{1}_{\{h=0\}}|f| \end{cases} \tag{22}$$

at $h \in \ell^\infty(\mathcal{X})$. The assertion then follows by the chain rule given in Proposition 3.6 in Shapiro (1990).

For a proof of (22), let (x_n) be a sequence in $\ell^\infty(\mathcal{X})$ converging to x and (t_n) be a sequence of positive real numbers converging to zero. We show that

$$\left\| \frac{\Phi_2(h + t_n x_n) - \Phi_2(h)}{t_n} - (\Phi_2)'_h(x) \right\|_1 \xrightarrow{n\to\infty} 0, \tag{23}$$

where

$$Z_n = \frac{\Phi_2(h + t_n x_n) - \Phi_2(h)}{t_n} - (\Phi_2)'_h(x)\ .$$

This proves the claim.

For a proof of (23), note that this statement is equivalent to

(A2.1) $\quad Z_n \xrightarrow{\lambda} 0\ ,$

(A2.2) $\quad (Z_n)$ is uniformly integrable ,

where $\xrightarrow{\lambda}$ denotes λ-stochastic convergence (see Theorem 21.4 and the preceding definitions in Bauer, 2011).

Proof of (A2.1) To prove this statement, it suffices to show every subsequence (Z_{n_k}) of (Z_n) has a further subsequence $(Z_{n_{k_j}})$, which converges to zero almost everywhere. So, let (Z_{n_k}) be a subsequence of (Z_n). Since $x_n \xrightarrow{\|\cdot\|_1} x$ by assumption, we know $x_{n_k} \xrightarrow{\|\cdot\|_1} x$. Theorem 15.7 in Bauer (2011) then implies that there exists a subsequence $(x_{n_{k_j}})$ such that $x_{n_{k_j}} \xrightarrow{a.e} x$. We conclude that $Z_{n_{k_j}} \xrightarrow{a.e} 0$ by the following:

1) On the set $\{t \in \mathbb{R} \mid x_{n_{k_j}}(t) \to x(t),\ h(t) = 0\}$, we have

$$Z_{n_{k_j}}(t) = \left|x_{n_{k_j}}(t)\right| - |x(t)| \xrightarrow{j\to\infty} 0.$$

2) On the set $\{t \in \mathbb{R} \mid x_{n_{k_j}}(t) \to x(t),\ h(t) > 0\}$, we have for sufficiently large j

$$Z_{n_{k_j}}(t) = x_{n_{k_j}}(t) - x(t) \xrightarrow{j\to\infty} 0.$$

3) On the set $\{t \in \mathbb{R} \mid x_{n_{k_j}}(t) \to x(t),\ h(t) < 0\}$, we have for sufficiently large j

$$Z_{n_{k_j}}(t) = -x_{n_{k_j}}(t) + x(t) \xrightarrow{j\to\infty} 0.$$

□

Proof of (A2.2) We have that

$$\begin{aligned}|Z_n| &= \left|\frac{|h + t_n x_n| - |h|}{t_n} - (\Phi_2)'_h(x)\right| \leq \left|\frac{|h + t_n x_n| - |h|}{t_n}\right| + \left|(\Phi_2)'_h(x)\right| \\ &\leq \frac{|h + t_n x_n - h|}{t_n} + |x| = |x_n| + |x|\end{aligned}$$

where we have used the definition of $(\Phi_2)'_h$ for the second inequality. Now $|x_n|$ is uniformly integrable, since $x_n \xrightarrow{\|\cdot\|_1} x$ and $|x|$ is uniformly integrable, since x is bounded. Therefore, $|x_n| + |x|$ is uniformly integrable. Since $|Z_n|$ is dominated by $|x_n| + |x|$, the result follows. □

A.2 Proof of Theorem 2

We first show that, unconditionally,

$$\hat{T} \xrightarrow{\mathbb{P}} T\ . \tag{24}$$

For this purpose, we note that from Assumption 1-7

$$\|\theta - \hat{\theta}\|_\infty = O_{\mathbb{P}}(n^{-1/2}). \tag{25}$$

Next, we define

$$S := \int_{\mathcal{N}^c} \operatorname{sgn}(\theta(x))\, \hat{G}(x)\, \mathrm{d}x + \int_{\mathcal{N}} |\hat{G}(x)|\, \mathrm{d}x$$

and let λ denote the Lebesgue measure on $\mathcal{X}$. By the triangle inequality, we have

$$|T - \hat{T}| \leq |T - S| + |S - \hat{T}|.$$

Note that $\{\hat{G}(x) - G(x)\}_{x\in\mathcal{X}}$ tends to $0 \in (\ell^\infty(\mathcal{X}), \|\cdot\|_\infty)$ in probability, which follows from the consistency of the estimators involved in the definition of $\hat{G}$ and the fact that m_l are twice continuously differentiable functions defined on a compact set. As a consequence of

$$|S - T| \le \|\hat{G} - G\|_1 \le \lambda(\mathcal{X})\|\hat{G} - G\|_\infty ,$$

we obtain $|T - S| \to 0$ in probability. We are hence left with showing that $|S - \hat{T}| \to 0$ in probability. To this end, we observe

$$|S - \hat{T}| \le A + B,$$

where

$$A = \Big| \int_{\hat{\mathcal{N}}^c} \operatorname{sgn}(\hat{\theta}(x))\hat{G}(x)\, \mathrm{d}x \; - \int_{\mathcal{N}^c} \operatorname{sgn}(\theta(x))\hat{G}(x)\, \mathrm{d}x \Big|,$$
$$B = \Big| \int_{\hat{\mathcal{N}}} |\hat{G}(x)|\, \mathrm{d}x \; - \int_{\mathcal{N}} |\hat{G}(x)|\, \mathrm{d}x \Big|.$$

We will only show that $A \to 0$ in probability; the corresponding result for B follows by similar arguments. Note that

$$\begin{aligned} A \le& \Big| \int_{\hat{\mathcal{N}}^c \cap \mathcal{N}} \operatorname{sgn}(\hat{\theta}(x))\hat{G}(x)\, \mathrm{d}x - \int_{\mathcal{N}^c \cap \hat{\mathcal{N}}} \operatorname{sgn}(\theta(x))\hat{G}(x)\, \mathrm{d}x \Big| \\ &+ \Big| \int_{\mathcal{N}^c \cap \hat{\mathcal{N}}^c} \big(\operatorname{sgn}(\hat{\theta}(x)) - \operatorname{sgn}(\theta(x))\big) \hat{G}(x) \mathrm{d}x \Big| \\ \le& \, \lambda(\hat{\mathcal{N}}^c \cap \mathcal{N})\|\hat{G}\|_\infty + \lambda(\mathcal{N}^c \cap \hat{\mathcal{N}})\|\hat{G}\|_\infty + o_{\mathbb{P}}(1) , \end{aligned}$$

where the last inequality is true because with high probability the signs in the third integral cancel each other out on $\hat{\mathcal{N}}^c \cap \mathcal{N}^c$. This can be seen by recalling the definition of the set $\mathcal{N}^c$ and (25). The other two terms vanish due to $\hat{G}(x)$ being bounded in probability by virtue of its tightness and because

$$\begin{aligned} \lambda(\hat{\mathcal{N}}^c \cap \mathcal{N}) &= \lambda\big(\{x \mid \hat{\theta}(x) \ge c\sqrt{\log(n)/n}, \theta(x) = 0\}\big) \\ &\le \lambda\big(\{x \mid \sqrt{n}(\hat{\theta}(x) - \theta(x)) \ge c \log(n)\}\big) = o_{\mathbb{P}}(1) \end{aligned}$$

due to (25). As a similar inequality holds true for the set $\mathcal{N}^c \cap \hat{\mathcal{N}}$, this concludes the proof of (24).

Define $\mathcal{Y} := \{Y_{\ell,i,j} : \ell = 1, 2,\ i = 1, \dots, k_\ell,\ j = 1, \dots, n_{\ell,i}\}$. Consider the conditional distribution $\mathbb{P}^{\hat{T}|\mathcal{Y}}$ and note that by the previous argument we have

$$\mathbb{P}(\hat{T} - T \in \mathcal{A}) = \int \mathbb{P}^{\hat{T}-T|\mathcal{Y}}(\mathcal{A}) d\mathbb{P} \to 0,$$

which implies that $\hat{T} - T|\mathcal{Y} \to 0$ in probability by a suitable choice of a countable family of $\mathcal{A}$ and a repeated subsequence argument. We see that $\hat{q}_{0,\alpha}$ converges to q_α in probability. As all quantities of which we take limits in the following are real-valued, we may assume, without loss of generality, that it does so even almost surely.

Observe that

$$\mathbb{P}\Big(d_1 \in [0, \hat{d}_1 - \frac{\hat{q}_{0,\alpha}}{\sqrt{n}}]\Big) = 1 - \mathbb{P}\Big(\sqrt{n}(\hat{d}_1 - d_1) \le \hat{q}_{0,\alpha}\Big)$$
$$= 1 - \mathbb{P}\Big(\sqrt{n}(\hat{d}_1 - d_1) \le q_\alpha + o(1)\Big).$$

By Egorov's theorem, we may assume that the $o(1)$ term vanishes uniformly on a set of measure δ for any $\delta > 0$. To be precise, for $n(m)$ large enough we have $o(1) \le 1/m$ on a set $\mathcal{A}_m$ that has measure at least $1 - 1/m$. Hence, for $n \ge n(m)$ we obtain

$$\mathbb{P}\Big(\sqrt{n}(\hat{d}_1 - d_1) \le q_\alpha + o(1)\Big) \le \mathbb{P}\Big(\sqrt{n}(\hat{d}_1 - d_1) \le q_\alpha + 1/m, \mathcal{A}_m\Big) + \mathbb{P}(\mathcal{A}_m^c)$$
$$\le \mathbb{P}\Big(\sqrt{n}(\hat{d}_1 - d_1) \le q_\alpha + 1/m\Big) + 1/m.$$

A similar lower bound can be obtained by the same arguments. Letting n go to infinity then establishes

$$\mathbb{P}\Big(d_1 \in [0, \hat{d}_1 - \frac{\hat{q}_{0,\alpha}}{\sqrt{n}}]\Big) = 1 - \mathbb{P}\Big(\sqrt{n}(\hat{d}_1 - d_1) \le \hat{q}_{0,\alpha}\Big) \to 1 - \alpha,$$

because the convergence of the distribution functions of $\sqrt{n}(\hat{d}_1 - d_1)$ is uniform for all continuity points of F_T.

This proves the first part of Theorem 2.

For the test in (12), under the null hypothesis $H_0 : d_1 \ge \epsilon$, we have $\epsilon - d_1 \le 0$, which implies for the probability of rejection

$$\mathbb{P}\Big(\hat{d}_1 < \epsilon + \frac{\hat{q}_{0,\alpha}}{\sqrt{n}}\Big) = \mathbb{P}(\sqrt{n}(\hat{d}_1 - d_1) < \sqrt{n}(\epsilon - d_1) + \hat{q}_{0,\alpha})$$
$$\le \mathbb{P}(\sqrt{n}(\hat{d}_1 - d_1) < \hat{q}_{0,\alpha})$$
$$= \mathbb{P}(\sqrt{n}(\hat{d}_1 - d_1) < q_\alpha - (q_\alpha - \hat{q}_{0,\alpha})) \xrightarrow{n\to\infty} \alpha,$$

where the convergence follows from similar arguments as for the first part of the theorem and Theorem 1. Consequently, the decision rule (12) defines an asymptotic level α -test. Similarly, under the alternative, we have $\epsilon - d_1 > 0$, which yields consistency, i.e.

$$\mathbb{P}\Big(\hat{d}_1 < \epsilon + \frac{\hat{q}_{0,\alpha}}{\sqrt{n}}\Big) = \mathbb{P}(\sqrt{n}(\hat{d}_1 - d_1) < \sqrt{n}(\epsilon - d_1) + \hat{q}_{0,\alpha}) \xrightarrow{n\to\infty} 1,$$

since $\hat{q}_{0,\alpha} \xrightarrow{\mathbb{P}} q_\alpha$ and $\sqrt{n}(\epsilon - d_1) \xrightarrow{n\to\infty} \infty$ imply $\sqrt{n}(\epsilon - d_1) + \hat{q}_{0,\alpha} \xrightarrow{n\to\infty} \infty$ and we know that $\sqrt{n}(\hat{d}_1 - d_1)$ converges in distribution by Theorem 1. □

A.3 Proof of Theorem 3

We start with proving the properties of the test. We have

$$\mathbb{P}\big(\hat{q}^*_{1-\alpha,0} < \epsilon\big) = \mathbb{P}\big(\sqrt{n}(\hat{q}^*_{1-\alpha,0} - \hat{d}_1) < \sqrt{n}(\epsilon - d_1) + \sqrt{n}(d_1 - \hat{d}_1)\big).$$

Following the arguments in the proof of Theorem 2 of Dette et al. (2018) (where we use Theorem 23.9 from Van der Vaart (2000) instead of an explicit first-order expansion and the continuous mapping theorem), we obtain

$$\sqrt{n}(\hat{q}^*_{1-\alpha,0} - \hat{d}_1) \xrightarrow{\mathbb{P}} q_{1-\alpha}, \tag{26}$$

where $q_{1-\alpha}$ is the $1-\alpha$ quantile of the random variable T defined in (7). Since $\sqrt{n}(d_1 - \hat{d}_1)$ converges in distribution to T by Theorem 1, T is symmetric when $\lambda(\mathcal{N}) = 0$, $\sqrt{n}(\epsilon - d_1)$ converges to zero if $d_1 = \epsilon$ and to $\pm\infty$ in the alternative/remainder of the null hypothesis, we obtain the desired statement on the significance level and the consistency of the test.

For the confidence interval, we observe that

$$\mathbb{P}\big(d_1 \in \hat{I}^*_n\big) = \mathbb{P}\big(d_1 < \hat{q}^*_{1-\alpha,0}\big) = \mathbb{P}\big(\sqrt{n}(d_1 - \hat{d}_1) < \sqrt{n}(\hat{q}^*_{1-\alpha,0} - \hat{d}_1)\big),$$

which yields the desired statement by (26). □

A.4 Proof of Theorem 4

Proof of (i) First, we determine the asymptotic distribution of the bootstrap test statistic $\hat{d}^*_1$. Define $\hat{\theta}^*(x) = m_1(x, \hat{\beta}^*_1) - m_2(x, \hat{\beta}^*_2)$ and $\hat{\hat{\theta}}(x) = m_1(x, \hat{\hat{\beta}}_1) - m_2(x, \hat{\hat{\beta}}_2)$. Following the proof of Theorem 1 in Dette et al. (2018) shows that conditionally on $\mathcal{Y}$ in probability

$$\big\{\sqrt{n}\big(\hat{\theta}^*(x) - \hat{\hat{\theta}}(x)\big)\big\}_{x\in\mathcal{X}} \xrightarrow{d} \{G(x)\}_{x\in\mathcal{X}}.$$

By assumption, the directional Hadamard derivative Φ'_θ is linear and thus a proper Hadamard derivative which allows us to apply the delta method for the bootstrap as stated in Theorem 23.9 in Van der Vaart (2000). Consequently we obtain

$$\sqrt{n}\big(\hat{d}^*_1 - \hat{\hat{d}}_1\big) = \sqrt{n}\big(\Phi(\{\hat{\theta}^*(x)\}_{x\in\mathcal{X}}) - \Phi(\{\hat{\hat{\theta}}(x)\}_{x\in\mathcal{X}})\big) \xrightarrow{d} \Phi'_\theta(\{G(x)\}_{x\in\mathcal{X}})$$

conditionally on $\mathcal{Y}$ in probability.

<u>*Case* 1 : $d_1 > \epsilon$.</u> We observe that

$$\begin{aligned}
\mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1}) =& \mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1},\ \hat{d}_1 \geq \epsilon) + \mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1},\ \hat{d}_1 < \epsilon) \\
\leq& \mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1},\ \hat{\hat{d}}_1 = \hat{d}_1) + \mathbb{P}(\hat{d}_1 < \epsilon) \\
\leq& \mathbb{P}(\hat{\hat{d}}_1 < \hat{q}^*_{\alpha,1}) + \mathbb{P}(\sqrt{n}(\hat{d}_1 - d_1) < \sqrt{n}(\epsilon - d_1)).
\end{aligned} \tag{27}$$

We now show that the first sequence in the upper bound (27) converges to zero. To prove this, first note that for all $\alpha \in (0, 1)$

$$\sqrt{n}\,(\hat{q}^*_{\alpha,1} - \hat{\hat{d}}_1) \xrightarrow{\mathbb{P}} q_\alpha, \tag{28}$$

where q_α denotes the α-quantile of the random variable T defined in (7). To see this, observe that

$$\alpha = \mathbb{P}(\hat{d}^*_1 < \hat{q}^*_{\alpha,1} \mid \mathcal{Y}) = \mathbb{P}(\sqrt{n}(\hat{d}^*_1 - \hat{\hat{d}}_1) < \sqrt{n}(\hat{q}^*_{\alpha,1} - \hat{\hat{d}}_1) \mid \mathcal{Y}) \text{ a.s .}$$

Since $\sqrt{n}(\hat{d}^*_1 - \hat{\hat{d}}_1)$ converges in distribution to T conditionally on $\mathcal{Y}$ in probability, Lemma 21.2 in Van der Vaart (2000) yields (28). Using (28) and choosing $\alpha > 0$ small enough such that $q_\alpha < 0$, we obtain

$$\begin{aligned}\mathbb{P}(\hat{\hat{d}}_1 < \hat{q}^*_{\alpha,1}) =& \mathbb{P}(\sqrt{n}(\hat{q}^*_{\alpha,1} - \hat{\hat{d}}_1) > 0)\\ &\le \mathbb{P}\big(\big|\sqrt{n}(\hat{q}^*_{\alpha,1} - \hat{\hat{d}}_1) - q_\alpha\big| > -q_\alpha\big) \xrightarrow{n\to\infty} 0.\end{aligned}$$

Finally, we show that the second sequence in the upper bound (27) converges to zero. Since $d_1 > \epsilon$ by assumption, we have that $\sqrt{n}(\epsilon - d_1) \to -\infty$, and from Theorem 1, we know that $\sqrt{n}(\hat{d}_1 - d_1)$ converges in distribution. Therefore, the result follows. This concludes the proof of (i) in the case $d_1 > \epsilon$.

<u>*Case* 2 : $d_1 = \epsilon$</u>. We observe that

$$\begin{aligned}\mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1}) =& \mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1},\ \hat{d}_1 \ge \epsilon) + \mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1},\ \hat{d}_1 < \epsilon)\\ =& \mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1},\ \hat{\hat{d}}_1 = \hat{d}_1) + \mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1},\ \hat{\hat{d}}_1 = \epsilon)\\ &- \mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1},\ \hat{d}_1 = \epsilon)\\ =& \mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1},\ \hat{\hat{d}}_1 = \hat{d}_1) + \mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1},\ \hat{\hat{d}}_1 = \epsilon = d_1) + o(1)\\ =& \mathbb{P}(\sqrt{n}(\hat{d}_1 - d_1) < \sqrt{n}(\hat{q}^*_{\alpha,1} - \hat{\hat{d}}_1),\ \hat{\hat{d}}_1 = \epsilon) + o(1)\\ =& \mathbb{P}(\sqrt{n}(\hat{d}_1 - d_1) < \sqrt{n}(\hat{q}^*_{\alpha,1} - \hat{\hat{d}}_1))\\ &- \mathbb{P}(\hat{d}_1 - d_1 < \hat{q}^*_{\alpha,1} - \hat{\hat{d}}_1, \hat{\hat{d}}_1 > \epsilon) + o(1).\end{aligned}$$

Because of (28) and Theorem 1, we have that

$$\mathbb{P}(\sqrt{n}(\hat{d}_1 - d_1) < \sqrt{n}(\hat{q}^*_{\alpha,1} - \hat{\hat{d}}_1)) \xrightarrow{n\to\infty} \alpha.$$

Since $\hat{\hat{d}}_1 > \epsilon$ implies $\hat{d}_1 - d_1 > 0$ and (28) holds, we obtain

$$\mathbb{P}(\hat{d}_1 - d_1 < \hat{q}^*_{\alpha,1} - \hat{\hat{d}}_1,\ \hat{\hat{d}}_1 > \epsilon) \le \mathbb{P}(0 < \hat{q}^*_{\alpha,1} - \hat{\hat{d}}_1) \xrightarrow{n\to\infty} 0,$$

which completes the proof of (i). ☐

Proof of (ii) The result follows by the same arguments as given for the proof of the second statement of Theorem 2 in Dette et al. (2018). Only note that the map $(b_1, b_2) \mapsto d_1(b_1, b_2)$ from $B_1 \times B_2$ onto $\mathbb{R}$ is uniformly continuous, since it is a continuous function on a compact set. ☐

A.5 Proof of the statement in Remark 4(ii)

Consider first the null hypothesis $H_0 : d_1 \geq \epsilon$. From Lemma 1 and Theorem 3.2 in Fang and Santos (2019), we know that conditionally in probability

$$\sqrt{n}\hat{\Phi}'^* := \hat{\Phi}'(\{\sqrt{n}(\hat{\theta}^* - \hat{\theta})\}_{x\in\mathcal{X}}) \xrightarrow{d} \Phi'_\theta(\{G(x)\}_{x\in\mathcal{X}}) = T.$$

<u>*Case* 1 : $d_1 > \epsilon$.</u> First note that for all $\alpha \in (0, 1)$ we have

$$\sqrt{n}\,\hat{q}^*_{\alpha,1} \xrightarrow{\mathbb{P}} q_\alpha, \tag{29}$$

where q_α denotes the α-quantile of T and $\hat{q}^*_{\alpha,1}$ denotes the α-quantile of $\hat{\Phi}'^*$. To see this, observe that by definition of $\hat{q}^*_{\alpha,1}$ we have

$$\alpha = \mathbb{P}(\hat{\Phi}'^* < \hat{q}^*_{\alpha,1} \mid \mathcal{Y}) = \mathbb{P}(\sqrt{n}\hat{\Phi}'^* < \sqrt{n}\hat{q}^*_{\alpha,1} \mid \mathcal{Y}) \text{ a.s.}$$

Since $\sqrt{n}\hat{\Phi}'^*$ converges in distribution to T conditionally on $\mathcal{Y}$ in probability, Lemma 21.2 in Van der Vaart (2000) yields (29). Using (29), we see that $\sqrt{n}(\hat{q}^*_{\alpha,1} + \epsilon - d_1) \xrightarrow{n\to\infty} -\infty$, since $d_1 > \epsilon$. Combining this result with the fact that $\sqrt{n}(\hat{d}_1 - d_1)$ converges in distribution by Theorem 1, we can conclude that

$$\mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1} + \epsilon) = \mathbb{P}(\sqrt{n}(\hat{d}_1 - d_1) < \sqrt{n}(\hat{q}^*_{\alpha,1} + \epsilon - d_1)) \xrightarrow{n\to\infty} 0.$$

<u>*Case* 2 : $d_1 = \epsilon$.</u> Since $\sqrt{n}(\hat{d}_1 - d_1)$ converges in distribution to T and (29) holds, we deduce that

$$\begin{aligned}\mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1} + \epsilon) =& \mathbb{P}(\sqrt{n}(\hat{d}_1 - d_1) < \sqrt{n}(\hat{q}^*_{\alpha,1} + \epsilon - d_1)) \\ =& \mathbb{P}(\sqrt{n}(\hat{d}_1 - d_1) < \sqrt{n}\hat{q}^*_{\alpha,1}) \xrightarrow{n\to\infty} \alpha.\end{aligned}$$

Next we consider the alternative $H_1 : d_1 < \epsilon$. Using (29) and $d_1 < \epsilon$, we deduce that $\sqrt{n}(\hat{q}^*_{\alpha,1} + \epsilon - d_1) \xrightarrow{n\to\infty} \infty$. Since $\sqrt{n}(\hat{d}_1 - d_1)$ converges in distribution, this implies that

$$\mathbb{P}(\hat{d}_1 < \hat{q}^*_{\alpha,1} + \epsilon) = \mathbb{P}(\sqrt{n}(\hat{d}_1 - d_1) < \sqrt{n}(\hat{q}^*_{\alpha,1} + \epsilon - d_1)) \xrightarrow{n\to\infty} 1.$$

☐

Lemma 1 *The sequence of functions*

$$\hat{\Phi}'(h) := \int_{|\hat{\theta}|\geq 1/s_n} \operatorname{sgn}(\hat{\theta}(x))h(x)\,\mathrm{d}x + \int_{|\hat{\theta}|<1/s_n} |h(x)|\,\mathrm{d}x$$

with $s_n/\sqrt{n} \to 0$ *satisfies Assumption* 4 *in Fang and Santos* (2019), *i.e. for* $h \in \ell^\infty(\mathcal{X})$ *we have*

$$\left|\hat{\Phi}'(h) - \Phi'_\theta(h)\right| \overset{\mathbb{P}}{\longrightarrow} 0$$

[*note that since* $\hat{\Phi}'$ *is Lipschitz continuous with respect to* $\|\cdot\|_1$ *it suffices to prove this simpler condition*; *see Fang and Santos* (2019)].

Proof Defining

$$\begin{aligned}
A &:= \left|\int_{|\hat{\theta}|\geq 1/s_n} \operatorname{sgn}(\hat{\theta}(x))h(x)\,\mathrm{d}x - \int_{|\theta|>0} \operatorname{sgn}(\theta(x))h(x)\,\mathrm{d}x\right|,\\
B &:= \left|\int_{|\hat{\theta}|<1/s_n} |h(x)|\,\mathrm{d}x - \int_{|\theta|=0} |h(x)|\,\mathrm{d}x\right|,
\end{aligned}$$

we note that $\left|\hat{\Phi}'(h) - \Phi'_\theta(h)\right| \leq A + B$, by the triangle inequality. Therefore, it suffices to show that $A \overset{\mathbb{P}}{\longrightarrow} 0$ and $B \overset{\mathbb{P}}{\longrightarrow} 0$. In order to show the former (the latter can be proven by similar arguments), we define the sets

$$M_1 := \left\{|\hat{\theta}| > \frac{1}{s_n}\right\}, \quad M_2 := \left\{|\theta| > 0\right\}$$

and note that

$$\begin{aligned}
A \leq\ & \left|\int_{M_1\cap M_2^c} \operatorname{sgn}(\hat{\theta}(x))h(x)\,\mathrm{d}x - \int_{M_2\cap M_1^c} \operatorname{sgn}(\theta(x))h(x)\,\mathrm{d}x\right|\\
&+ \left|\int_{M_1\cap M_2} \Big(\operatorname{sgn}(\hat{\theta}(x)) - \operatorname{sgn}(\theta(x))\Big)h(x)\,\mathrm{d}x\right|\\
\leq\ & \lambda(M_1\cap M_2^c)\|h\|_\infty + \lambda(M_2\cap M_1^c)\|h\|_\infty + o_{\mathbb{P}}(1),
\end{aligned} \tag{30}$$

due to $\hat{\theta} \overset{\mathbb{P}}{\longrightarrow} \theta$ and where λ denotes the Lebesgue measure. Therefore, it suffices to show that the first two summands in (30) converge to zero in probability. Regarding the first term, we have

$$\begin{aligned}
\lambda\Big(M_1\cap M_2^c\Big) &= \lambda\Big(|\hat{\theta}| > \frac{1}{s_n}, \theta = 0\Big) = \lambda\Big(s_n|\hat{\theta} - \theta| > 1, \theta = 0\Big)\\
&\leq \lambda\Big(s_n|\hat{\theta} - \theta| > 1\Big) = \lambda\Big(\frac{s_n}{\sqrt{n}}\sqrt{n}|\hat{\theta} - \theta| > 1\Big),
\end{aligned}$$

where the last term converges to zero, since $s_n/\sqrt{n} \to 0$ by assumption and since the sequence $\sqrt{n}|\hat{\theta} - \theta|$ is tight. The second summand can be handled similarly. □

References

Aitchison, J. (1964). Confidence-region tests. *Journal of the Royal Statistical Society, Series B, 26*, 462–476.

Bauer, H. (2011). *Measure and integration theory*, Vol. 26. New York: Walter de Gruyter.

Berger, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics, 24*, 295–300.

Bradley, A. P. (1997). The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition, 30*(7), 1145–1159.

Chow, S.-C., Liu, P.-J. (1992). *Design and analysis of bioavailability and bioequivalence studies*. New York: Marcel Dekker.

Cox, T., Czanner, G. (2016). A practical divergence measure for survival distributions that can be estimated from Kaplan–Meier Curves. *Statistics in Medicine, 35*, 66.

Dette, H., Möllenhoff, K, Volgushev, S, Bretz, F. (2018). Equivalence of regression curves. *Journal of the American Statistical Association, 113*(522), 711–729.

EMA. (2014). *Guideline on the investigation of bioequivalence*. European Medicines Agency. Available at http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/01/WC500070039.pdf

Fang, Z., Santos, A. (2019). Inference on directionally differentiable functions. *The Review of Economic Studies, 86*(1), 377–412.

Gsteiger, S., Bretz, F., Liu, W. (2011). Simultaneous confidence bands for nonlinear regression models with application to population pharmacokinetic analyses. *Journal of Biopharmaceutical Statistics, 21*(4), 708–725.

Hauschke, D., Steinijans, V., Pigeot, I. (2007). *Bioequivalence studies in drug development methods and applications. Statistics in practice*. New York: Wiley.

Heller, G., Seshan, V. E., Moskowitz, C. S., Gönen, M. (2016). Inference for the difference in the area under the ROC curve derived from nested binary regression models. *Biostatistics, 18*(2), 260–274.

Jachno, K., Heritier, S., Wolfe, R. (2019). Are non-constant rates and non-proportional treatment effects accounted for in the design and analysis of randomised controlled trials? A review of current practice. *BMC Medical Research Methodology, 19*(1), 1–9.

Liu, W., Hayter, A. J., Wynn, H. P. (2007). Operability region equivalence: Simultaneous confidence bands for the equivalence of two regression models over restricted regions. *Biometrical Journal, 49*(1), 144–150.

Liu, W., Bretz, F., Hayter, A. J., Wynn, H. P. (2009). Assessing non-superiority, non-inferiority of equivalence when comparing two regression models over a restricted covariate region. *Biometrics, 65*(4), 1279–1287.

McCaw, Z.R., Yin, G., Wei, L.-J. (2019). Using the restricted mean survival time difference as an alternative to the hazard ratio for analyzing clinical cardiovascular studies. *Circulation, 140*(17), 1366–1368.

Möllenhoff, K., Dette, H., Bretz, F. (2022). Testing for similarity of binary efficacy-toxicity responses. *Biostatistics, 23*(3), 949–966 .

Ostrovski, V. (2017). Testing equivalence of multinomial distributions. *Statistics & Probability Letters, 124*, 77–82.

Pepe, M. S., Kerr, K. F., Longton, G., Wang, Z. (2013). Testing for improvement in prediction model performance. *Statistics in Medicine, 32*(9), 1467–1482.

Royston, P., Parmar, M. K. (2013). Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Medical Research Methodology, 13*(1), 152.

Shapiro, A. (1990). On concepts of directional differentiability. *Journal of optimization theory and applications, 66*(3), 477–487.

Shapiro, A. (1991). Asymptotic analysis of stochastic programs. *Annals of Operations Research, 30*(1), 169–186.

U.S. Food and Drug Administration. (2003). *Guidance for industry: Bioavailability and bioequivalence studies for orally administered drug products-general considerations*. Washington, DC: Food and Drug Administration. Available at https://www.ipqpubs.com/wp-content/uploads/2020/12/BioStudies_OralDosageProducts_March.2003.GUIDANCE.pdf.pdf

Van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge: Cambridge University Press.