Equivalence of dose response curves

Holger Dette, Kathrin Möllenhoff,
Stanislav Volgushev
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: holger.dette@rub.de

Frank Bretz
Novartis Pharma AG
Lichtstrasse 35
4002 Basel, Switzerland
e-mail: frank.bretz@novartis.com

May 19, 2015

Abstract

This paper investigates the problem if the difference between two parametric models m_1, m_2 describing the relation between the response and covariates in two groups is of no practical significance, such that inference can be performed on the basis of the pooled sample. Statistical methodology is developed to test the hypotheses $H_0 : d(m_1, m_2) \geq \varepsilon$ versus $H_1 : d(m_1, m_2) < \varepsilon$ of equivalence between the two regression curves m_1, m_2, where d denotes a metric measuring the distance between m_1 and m_2 and ε is a pre specified constant. Our approach is based on an estimate $d(\hat{m}_1, \hat{m}_2)$ of this distance and its asymptotic properties. In order to improve the approximation of the nominal level for small sample sizes a bootstrap test is developed, which addresses the specific form of the interval hypotheses. In particular, data has to be generated under the null hypothesis, which implicitly defines a manifold for the vector of parameters. The results are illustrated by means of a simulation study, and it is demonstrated that the new methods yield a substantial improvement with respect to power compared to all currently available tests for this problem.

Keywords and Phrases: dose response studies; nonlinear regression; equivalence of curves; constrained parameter estimation; parametric bootstrap

1 Introduction

A frequent problem in statistics is the comparison of two regression models which are used for the description of the relation between the response variable and covariates for two different groups, respectively. An important aspect in this type of problems is to investigate whether the differences between the two models for the two groups are of no practical significance, so that only one model can be used for both groups. Such problems appear for example in population pharmacokinetics (PK) where the goal is to establish the PK bio-equivalence of the concentration versus time profiles, say m_1, m_2, of two compounds. "Classical" bio-equivalence methods usually establish equivalence between real valued quantities such as the area under the curve (AUC) or the maximum concentrations ($C_{\max}$) [see Chow and Liu (1992); Hauschke et al. (2007)]. However, such an approach may be misleading because the two profiles could be very different although they may have similar AUC or $C_{\max}$ values. From this perspective it might be more reasonable to work directly with the underlying PK profiles instead of summaries of this type.

The problem of establishing the equivalence of two regression models at a controlled type I error has found considerable attention in the recent literature. For example, Liu et al. (2009) proposed tests for the hypothesis of equivalence of two regression functions, which are applicable in linear models. Gsteiger et al. (2011) considered non-linear models and suggested a bootstrap method which is based on a confidence band for the difference of the two regression models. Both references use the unit intersection principle to construct the test. We will demonstrate in the present paper that this approach yields to a rather conservative method with very low power. As an alternative, we propose to estimate the distance, say $d(m_1, m_2)$, between the regression curves directly and to decide for the equivalence of the two curves if the estimator is smaller than a given threshold. The critical values of this test can be obtained by asymptotic theory, which describes the limiting distribution of an appropriately standardized estimated distance. In order to improve the approximation of the nominal level for small samples sizes a non-standard bootstrap approach is proposed to determine critical values of this test.

In Section 2 we introduce the general problem of equivalent regression curves. While the concept of similarity of the two profiles is formulated here for a general metric d, we concentrate in the subsequent discussion on two specific cases. Section 3 is devoted to the comparison of curves with respect to L^2-distances. Such distances are attractive for PK models because they measure the squared integral of the difference between the two curves and are therefore related to the area under the curve. We prove asymptotic normality of the corresponding test statistic and construct an asymptotic level α-test. Moreover, a new bootstrap procedure is introduced, which addresses the particular difficulties arising in the problem of testing interval hypotheses. In particular resampling has to be performed under the null hypothesis $H_0 : d(m_1, m_2) \geq \varepsilon$, which defines (implicitly) a manifold in the parameter space. We prove consistency of the bootstrap test and demonstrate by means of a simulation study that it yields to an improvement of the approximation of the nominal level for small sample sizes. In Section 4 the maximal deviation between the two curves is considered as a measure of similarity, for which corresponding results are substantially harder to derive. For example, we prove weak convergence of a corresponding test statistic, but the limit distribution depends in a complicated way on the extremal points of the "true" difference. This problem is again solved by developing a bootstrap test. The finite sample properties of the new methodology are illustrated in Section 5, where we also provide a comparison with the method of Gsteiger et al. (2011). In particular, it is demonstrated that the methodology proposed in this paper yields a substantially more powerful procedure than the test proposed by these authors. Finally, all technical details and proofs (which are complicated) are deferred to an appendix in Section 6.

2 Equivalence of regression curves

We use two regression models to describe the relationship between the response variables and covariates for the two different groups, that is

$$Y_{1,i,j} = m_1(x_{1,i}, \beta_1) + \varepsilon_{1,i,j}\ ,\ j = 1, \dots, n_{1,i},\ i = 1, \dots, k_1 \tag{2.1}$$

$$Y_{2,i,j} = m_2(x_{2,i}, \beta_2) + \varepsilon_{2,i,j}\ ,\ j = 1, \dots, n_{2,i},\ i = 1, \dots, k_2 \tag{2.2}$$

Here the covariate region is denoted by $\mathcal{X} \subset \mathbb{R}^d$, $x_{\ell,i}$ represents the ith dose level (in group ℓ), $n_{\ell,i}$ is the number of patients treated at dose level $x_{\ell,i}$ and k_ℓ denotes the number of different dose levels in group ℓ $(= 1, 2)$. The sample size in each group is denoted by $n_\ell = \sum_{i=1}^{k_\ell} n_{\ell,i}$ $(\ell = 1, 2)$, $n = n_1 + n_2$ is the total sample size, and in (2.1) and (2.2) the functions m_1 and m_2 define the (non-linear) regression models with p_1- and p_2-dimensional parameters β_1 and β_2, respectively. The error terms are assumed to be independent and identically distributed with mean 0 and variance σ_ℓ^2 for group ℓ $(= 1, 2)$. Let $(\mathcal{M}, d)$ denote a metric space of real valued functions of the form $g : \mathcal{X} \to \mathbb{R}$ with metric d. We assume (for all β_1, β_2) that the regression functions satisfy $m_1(\cdot, \beta_1), m_2(\cdot, \beta_2) \in \mathcal{M}$, identify the models m_ℓ by their parameters β_ℓ and denote the distance between the two models by $d(\beta_1, \beta_2)(= d(m_1, m_2))$.
We consider the curves m_1 and m_2 as *equivalent* if the distance between the two curves is small, that is $d(\beta_1, \beta_2) < \varepsilon$, where ε is a positive constant specified by the experimenter. In order to establish "equivalence" of the two dose response curves at a controlled type I error, we formulate the hypotheses

$$H_0 : d(\beta_1, \beta_2) \geq \varepsilon \quad \text{versus} \quad H_1 : d(\beta_1, \beta_2) < \varepsilon. \tag{2.3}$$

In the following sections we are particularly interested in the metric space of all continuous functions with metric

$$d_\infty(\beta_1, \beta_2) = \max_{x \in \mathcal{X}} |m_1(x, \beta_1) - m_2(x, \beta_2)| \tag{2.4}$$

and of all square integrable functions with metric

$$d_2(\beta_1, \beta_2) = \Big(\int_{\mathcal{X}} |m_1(x, \beta_1) - m_2(x, \beta_2)|^2 dx \Big)^{1/2}. \tag{2.5}$$

The choice of the distance d depends on the specific problem under consideration. The metric d_∞ is of interest in drug stability studies, where one investigates whether the maximum difference in mean drug content between two batches is no larger than a pre-specified threshold. The metric d_2 might be useful if the distance should also reflect important pharmaceutical quantities such as the area under the curve.

The metric (2.4) has also been considered in Liu et al. (2009) and Gsteiger et al. (2011), who constructed a confidence band for the difference of the two regression curves and used the intersection-union-test [see for example Berger (1982)] to derive a test for the hypothesis that the two curves are equivalent. In linear models (with normally distributed errors) this test keeps its level not only asymptotically but also for fixed sample size. However, the resulting test is extremely conservative, and as it will be demonstrated in Section 5, has a very low power. In the following discussion we will develop alternatives and substantially more powerful tests for the equivalence of the two regression curves. Roughly speaking, we consider for $\ell = 1, 2$ the estimate $m_\ell(\cdot, \hat\beta_\ell)$ of the regression curve m_ℓ and reject the null hypothesis in (2.3) for small values of the statistic $\hat d = d(\hat\beta_1, \hat\beta_2)$. The critical values can be obtained by asymptotic theory deriving the limiting distribution of $\sqrt{n}\,(\hat d - d)$ as $n_1, n_2 \to \infty$, which will be developed in the following sections. This approach leads to a satisfactory solution for the L^2-distance (2.5) based on the quantiles of the normal distribution (see Section 3). However, for the maximal deviation distance (2.4), the limiting distribution depends in a complicated way on the extremal points

$$\mathcal{E} = \{x \in \mathcal{X} \mid |\Delta(x, \beta_1, \beta_2)| = d_\infty(\beta_1, \beta_2)\}$$

of the "true" difference

$$\Delta(x, \beta_1, \beta_2) = m_1(x, \beta_1) - m_2(x, \beta_2). \tag{2.6}$$

Moreover, in small population trials the approximation of the nominal level of tests developed by asymptotic theory may not be very precise. In order to obtain a more accurate approximation of the nominal level a non-standard bootstrap procedure is proposed and its consistency is proved. This procedure has to be constructed in a way such that it addresses the particular features of the hypothesis of equivalence of the curves, which is defined in (2.3). In particular, data has to be generated under the null hypothesis $d(\beta_1, \beta_2) \geq \varepsilon$, which implicitly defines a manifold for the vector of parameters $(\beta_1^T, \beta_2^T)^T \in \mathbb{R}^{p_1+p_2}$ of both models. The non-differentiability of the metric d_∞ exhibits some technical difficulties of such an approach, and for this reason we begin the discussion with the L^2-distance d_2.

3 Comparing curves by L^2-distances

In this chapter we construct a test for the equivalence of the two regression curves with respect to the squared L^2-norm, i.e. we consider the hypotheses of the form

$$H_0 : \int_{\mathcal{X}} (m_1(x,\beta_1) - m_2(x,\beta_2))^2 dx \geq \varepsilon \quad \text{versus} \quad H_1 : \int_{\mathcal{X}} (m_1(x,\beta_1) - m_2(x,\beta_2))^2 dx < \varepsilon. \tag{3.1}$$

To be precise, note that under regularity assumptions [see Section 6 for details] the ordinary least squares (OLS) estimators, say $\hat\beta_1$ and $\hat\beta_2$, of the parameters β_1 and β_2 can usually be "linearized" in the form

$$\sqrt{n_\ell}\ (\hat\beta_\ell - \beta_\ell) = \frac{1}{\sqrt{n_\ell}} \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} \phi_{\ell,i,j} + o_{\mathbb{P}}(1), \qquad \ell = 1,2, \tag{3.2}$$

where the functions $\phi_{\ell,i,j}$ are given by

$$\phi_{\ell,i,j} = \tfrac{\varepsilon_{\ell,i,j}}{\sigma_\ell^2} \Sigma_\ell^{-1} \tfrac{\partial}{\partial b_\ell} m_\ell(x_{\ell,i,}, b_\ell)\big|_{b_\ell=\beta_\ell}, \qquad \ell = 1,2, \tag{3.3}$$

and the $p_1 \times p_1$ and $p_2 \times p_2$ dimensional matrices Σ_1 and Σ_2 are defined by

$$\Sigma_\ell = \frac{1}{\sigma_\ell^2} \sum_{i=1}^{k_\ell} \zeta_{\ell,i} \tfrac{\partial}{\partial b_\ell} m_\ell(x_{\ell,i,}, b_\ell)\big|_{b_\ell=\beta_\ell} \big(\tfrac{\partial}{\partial b_\ell} m_\ell(x_{\ell,i,}, b_\ell)\big|_{b_\ell=\beta_\ell}\big)^T, \quad \ell = 1,2. \tag{3.4}$$

For these arguments we assume (besides the regularity assumptions commonly made for OLS-estimation) that the matrices Σ_ℓ are non-singular and that the sample sizes n_1 and n_2 converge to infinity such that

$$\lim_{n_\ell \to \infty} \frac{n_{\ell,i}}{n_\ell} = \zeta_{\ell,i} > 0\ , \quad i = 1, \dots, k_\ell\ ,\ \ell = 1,2 \tag{3.5}$$

and

$$\lim_{n_1,n_2 \to \infty} \frac{n}{n_1} = \lambda \in (1, \infty). \tag{3.6}$$

It follows by a straightforward calculation that the ordinary least squares estimators are asymptotically normal distributed, i.e.

$$\sqrt{n_\ell}(\hat\beta_\ell - \beta_\ell) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_\ell^{-1})\ ,\ \ell = 1,2, \tag{3.7}$$

where the symbol $\xrightarrow{\mathcal{D}}$ means weak convergence (convergence in distribution for real valued random variables). The asymptotic variance in (3.7) can easily be estimated by replacing the parameters β_ℓ and $\zeta_{\ell,i}$ in (3.4) by their estimates $\hat\beta_\ell$, and $n_{\ell,i}/n_\ell$ ($\ell = 1,2$). The resulting estimator will be denoted by $\hat\Sigma_\ell$ throughout this paper. The null hypothesis in (3.1) is rejected, whenever the inequality

$$\hat d_2 := d_2(\hat\beta_1, \hat\beta_2) = \int_{\mathcal{X}} (m_1(x,\hat\beta_1) - m_2(x,\hat\beta_2))^2 dx < c \tag{3.8}$$

is satisfied, where c is a pre-specified constant which defines the level of the test. In order to determine this constant we will derive the asymptotic distribution of the statistic $\hat{d}_2$. The following result is proved in Appendix 6.1.

Theorem 3.1. *If Assumptions 6.1 - 6.5 from the Appendix,* (3.5) *and* (3.6) *are satisfied, we have*

$$\sqrt{n}(\hat{d}_2 - d_2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_{d_2}), \tag{3.9}$$

where the asymptotic variance is given by

$$\sigma^2_{d_2} = \sigma^2_{d_2}(\beta_1, \beta_2) = 4 \int_{\mathcal{X}\times\mathcal{X}} \Delta(x, \beta_1, \beta_2)\Delta(y, \beta_1, \beta_2)k(x, y)dxdy, \tag{3.10}$$

$\Delta(x, \beta_1, \beta_2)$ *and the kernel* $k(x, y)$ *are defined by* (2.6) *and*

$$k(x, y) := \lambda\big(\tfrac{\partial}{\partial\beta_1} m_1(x, \beta_1)\big)^T \Sigma_1^{-1} \tfrac{\partial}{\partial\beta_1} m_1(y, \beta_1) + \tfrac{\lambda}{\lambda-1}\big(\tfrac{\partial}{\partial\beta_2} m_2(x, \beta_2)\big)^T \Sigma_2^{-1} \tfrac{\partial}{\partial\beta_2} m_2(y, \beta_2), \tag{3.11}$$

respectively.

Theorem 3.1 provides an asymptotic level α test for the hypothesis (3.1) of equivalence of the two regression curves. More precisely, if $\hat{\sigma}^2_{d_2} = \sigma^2_{d_2}(\hat{\beta}_1, \hat{\beta}_2)$ denotes the (canonical) estimator of the asymptotic variance in (3.10), then the null hypothesis in (3.1) is rejected if

$$\hat{d}_2 < \varepsilon + \frac{\hat{\sigma}_{d_2}}{\sqrt{n}} u_\alpha, \tag{3.12}$$

where u_α is the α-quantile of the standard normal distribution. The finite sample properties of this test will be investigated in Section 5.1.

Remark 3.2. It follows from Theorem 3.1 that the test (3.12) has asymptotic level α and is consistent if $n_1, n_2 \to \infty$. More precisely, if Φ denotes the cumulative distribution function of the standard normal distribution we have for the probability of rejecting the null hypothesis in (3.1)

$$\mathbb{P}\Big(\hat{d}_2 < \varepsilon + \frac{\hat{\sigma}_{d_2}}{\sqrt{n}} u_\alpha\Big) = \mathbb{P}\Big(\frac{\sqrt{n}}{\hat{\sigma}_{d_2}}(\hat{d}_2 - d_2) < \frac{\sqrt{n}}{\hat{\sigma}_{d_2}}(\varepsilon - d_2) + u_\alpha\Big) .$$

Under continuity assumptions it follows that $\hat{\sigma}^2_{d_2} \xrightarrow{\mathbb{P}} \sigma^2_{d_2}$ and Theorem 3.1 yields $\sqrt{n}\ (\hat{d}_2 - d_2)/\hat{\sigma}_{d_2} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$. This gives

$$\mathbb{P}\Big(\hat{d}_2 \leq \varepsilon + \frac{\hat{\sigma}_{d_2}}{\sqrt{n}} u_\alpha\Big) \underset{n_1,n_2\to\infty}{\longrightarrow} \begin{cases} 0 \text{ if } d_2 > \varepsilon \\ \alpha \text{ if } d_2 = \varepsilon \\ 1 \text{ if } d_2 < \varepsilon \end{cases} .$$

The test (3.12) can be recommended if the sample sizes are reasonable large. However, we will demonstrate in Section 5 that for very small sample sizes, the critical values provided by this asymptotic theory may not provide an accurate approximation of the nominal level, and for this reason we will also investigate a parametric bootstrap procedure to generate critical values for the statistic $\hat{d}_2$.

Algorithm 3.3. (parametric bootstrap for testing precise hypotheses)

(1) Calculate the OLS-estimators $\hat{\beta}_1$ and $\hat{\beta}_2$, the corresponding variance estimators

$$\hat{\sigma}_\ell^2 = \frac{1}{n_\ell} \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} (Y_{\ell,i,j} - m_\ell(x_{\ell,i}, \hat{\beta}_\ell))^2; \qquad \ell = 1, 2,$$

and the test statistic $\hat{d}_2 = d_2(\hat{\beta}_1, \hat{\beta}_2)$ defined by (3.8).

(2) Define estimators of the parameters β_1 and β_2 by

$$\hat{\hat{\beta}}_\ell = \begin{cases} \hat{\beta}_\ell \text{ if } \hat{d}_2 \geq \varepsilon \\ \tilde{\beta}_\ell \text{ if } \hat{d}_2 < \varepsilon \end{cases} \qquad \ell = 1, 2, \tag{3.13}$$

where $\tilde{\beta}_1, \tilde{\beta}_2$ denote the OLS-estimators of the parameters β_1, β_2 under the constraint

$$d_2(\beta_1, \beta_2) = \int_{\mathcal{X}} (m_1(x, \beta_1) - m_2(x, \beta_2))^2 dx = \varepsilon. \tag{3.14}$$

Finally, define $\hat{\hat{d}}_2 = d_2(\hat{\hat{\beta}}_1, \hat{\hat{\beta}}_2)$ and note that $\hat{\hat{d}}_2 \geq \varepsilon$.

(3) Bootstrap test

(i) Generate bootstrap data under the null hypothesis, that is

$$Y^*_{\ell,i,j} = m_\ell(x_{\ell,i}, \hat{\hat{\beta}}_\ell) + \varepsilon^*_{\ell,i,j} \; , i = 1, \ldots, n_{\ell,i}, \;\; \ell = 1, 2, \tag{3.15}$$

where the errors $\varepsilon^*_{\ell,i,j}$ are independent normally distributed such that $\varepsilon^*_{\ell,i,j} \sim \mathcal{N}(0, \hat{\sigma}_\ell^2)$ $(\ell = 1, 2)$.

(ii) Calculate the OLS estimates $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ from the bootstrap data and the test statistic

$$\hat{d}_2^* = d_2(\hat{\beta}_1^*, \hat{\beta}_2^*) = \int_{\mathcal{X}} (m_1(x, \hat{\beta}_1^*) - m_2(x, \hat{\beta}_2^*))^2 dx$$

from the bootstrap data. The quantile of the distribution of the statistic $\hat{d}_2^*$ (which depends on the data $\{Y_{l,i,j} | l = 1, 2; \; j = 1, ...n_{l,i}; \; i = 1, ..., k_l\}$ through the estimates $\hat{\hat{\beta}}_1$ and $\hat{\hat{\beta}}_2$) is denoted by $\hat{q}_\alpha$.

The steps (i) and (ii) are repeated B times to generate replicates $\hat{d}^*_{2,1}, \ldots, \hat{d}^*_{2,B}$ of $\hat{d}^*_2$. If $\hat{d}_2^{*(1)} \leq \ldots \leq \hat{d}_2^{*(B)}$ denotes the corresponding order statistic, the estimator of the quantile of the distribution of $\hat{d}^*_2$ is defined by $\hat{q}_\alpha^{(B)} := \hat{d}_2^{*(\lfloor B\alpha \rfloor)}$, and the null hypothesis is rejected whenever

$$\hat{d}_2 < \hat{q}_\alpha^{(B)}. \tag{3.16}$$

The following result shows that the bootstrap test (3.16) is a consistent asymptotic level α-test.

Theorem 3.4. *Assume that the conditions of Theorem 3.1 are satisfied.*

(1) If the null hypothesis in (3.1) *holds, then we have for any* $\alpha \in (0, 0.5)$

$$\lim_{n_1,n_2\to\infty} \mathbb{P}(\hat{d}_2 < \hat{q}_\alpha) = \begin{cases} 0 & \text{if } d_2 > \varepsilon \\ \alpha & \text{if } d_2 = \varepsilon \end{cases}. \tag{3.17}$$

(2) If the alternative in (3.1) *holds, then we have for any* $\alpha \in (0, 0.5)$

$$\lim_{n_1,n_2\to\infty} \mathbb{P}(\hat{d}_2 < \hat{q}_\alpha) = 1. \tag{3.18}$$

4 Comparing curves by their maximal deviation

This section is devoted to a test for the hypotheses (2.3), where d denotes the maximal absolute deviation defined by (2.4). The corresponding test statistic is given by the maximum distance

$$\hat{d}_\infty = d_\infty(\hat{\beta}_1, \hat{\beta}_2) = \max_{x\in\mathcal{X}} |m_1(x, \hat{\beta}_1) - m_2(x, \hat{\beta}_2)| \tag{4.1}$$

between the two estimated regression functions, where $\hat{\beta}_1, \hat{\beta}_2$ are the OLS-estimates from the two samples. In order to describe the asymptotic distribution of the statistic $\hat{d}_\infty$ we define

$$\mathcal{E} = \big\{x \in \mathcal{X} \big|\ |m_1(x, \beta_1) - m_2(x, \beta_2)| = d_\infty\big\} \tag{4.2}$$

as the set of extremal points and introduce the decomposition $\mathcal{E} = \mathcal{E}^+ \cup \mathcal{E}^-$, where

$$\mathcal{E}^\mp = \big\{x \in \mathcal{X} \big|\ m_1(x, \beta_1) - m_2(x, \beta_2) = \mp\, d_\infty\big\}. \tag{4.3}$$

The following result is proved in Section 6.3 of the Appendix.

Theorem 4.1. *If* $d_\infty > 0$ *and the assumptions of Theorem 3.1 are satisfied, then*

$$\sqrt{n}\ (\hat{d}_\infty - d_\infty) \xrightarrow{\mathcal{D}} \mathcal{Z} := \max\big\{\max_{x\in\mathcal{E}^+} G(x), \max_{x\in\mathcal{E}^-}(-G(x))\big\}, \tag{4.4}$$

where $\{G(x)\}_{x\in\mathcal{X}}$ *denotes a Gaussian process defined by*

$$G(x) = \big(\tfrac{\partial}{\partial\beta_1} m_1(x, \beta_1)\big)^T \sqrt{\lambda}\Sigma_1^{-1/2} Z_1 - \big(\tfrac{\partial}{\partial\beta_2} m_2(x, \beta_2)\big)^T \sqrt{\tfrac{\lambda}{\lambda-1}}\Sigma_2^{-1/2} Z_2, \tag{4.5}$$

and Z_1 *and* Z_2 *are* p_1*- and* p_2*-dimensional standard normal distributed random variables, respectively, i.e.* $Z_\ell \sim \mathcal{N}(0, I_{p_\ell})$, $\ell = 1, 2$.

In principle, Theorem 4.1 provides an asymptotic level α-test for the hypotheses

$$H_0 : d_\infty(\beta_1, \beta_2) \geq \varepsilon \quad \text{versus} \quad H_1 : d_\infty(\beta_1, \beta_2) < \varepsilon \tag{4.6}$$

by rejecting the null hypotheses whenever $\hat{d}_\infty < q_{\alpha,\infty}$, where $q_{\alpha,\infty}$ denotes the α-quantile of the distribution of the random variable $\mathcal{Z}$ defined in (4.4). However, this distribution has a very complicated structure, which depends on the unknown location of the extremal points of the "true" difference $\Delta(\cdot, \beta_2, \beta_2) = m_1(\cdot, \beta_1) - m_2(\cdot, \beta_2)$. For example, if $\mathcal{E} = \{x_0\}$ the distribution of $\mathcal{Z}$ is a centered normal distribution but with variance

$$\sigma_\infty^2 = \lambda\big(\tfrac{\partial}{\partial\beta_1} m_1(x_0, \beta_1)\big)^T \Sigma_1^{-1} \tfrac{\partial}{\partial\beta_1} m_1(x_0, \beta_1) + \tfrac{\lambda}{\lambda-1}\big(\tfrac{\partial}{\partial\beta_2} m_2(x_0, \beta_2)\big)^T \Sigma_2^{-1} \tfrac{\partial}{\partial\beta_2} m_2(x_0, \beta_2) \tag{4.7}$$

which depends on the location of the (unique) extremal point x_0. In general (more precisely in the case $\#\mathcal{E} > 1$) the distribution of $\mathcal{Z}$ is the distribution of a maximum of dependent Gaussian random variables, where the variances and the dependence structure depend on the location of the extremal points of the function $\Delta(\cdot, \beta_1, \beta_2)$. Because the estimation of these points is very difficult, we propose a bootstrap approach to obtain quantiles. The bootstrap test is defined in the same way as described in Algorithm 3.3, where the distance d_2 is replaced by the maximal deviation d_∞. The corresponding quantile obtained in Step 3(ii) of Algorithm 3.3 is now denoted by $\hat{q}_{\alpha,\infty}$. The following result is proved in Section 6.4 of the Appendix and shows that the test, which rejects the null hypothesis in (4.9), whenever

$$\hat{d}_\infty < \hat{q}_{\alpha,\infty} \tag{4.8}$$

has asymptotic level α and is consistent if the cardinality of the set $\mathcal{E}$ is one.

Theorem 4.2. *If the assumptions of Theorem 4.1 are satisfied and the set $\mathcal{E}$ defined in* (4.2) *consists of one point, then the test* (4.8) *is a consistent asymptotic level α-test, that is*

(1) If the null hypothesis in (4.6) *is satisfied, then we have for any $\alpha \in (0, 0.5)$*

$$\lim_{n\to\infty} \mathbb{P}\big(\hat{d}_\infty < \hat{q}_{\alpha,\infty}\big) \leq \alpha, \tag{4.9}$$

(2) If the alternative in (4.6) *is satisfied, then we have for any $\alpha \in (0, 0.5)$*

$$\lim_{n\to\infty} \mathbb{P}\big(\hat{d}_\infty < \hat{q}_{\alpha,\infty}\big) = 1. \tag{4.10}$$

Note that the condition that the set $\mathcal{E}$ should only contain one point in Theorem 4.2 above is critical. If the set $\mathcal{E}$ contains more than one point, the corresponding bootstrap test will usually be conservative [see Section 5.2 for some numerical results]. Intuitively, this behavior can be explained by the fact that the limiting distribution in Theorem 4.1 depends on the parameters in a discontinuous way.

5 Finite sample properites

In this section we investigate the finite sample properties of the test proposed in Section 3. We study the power and the approximation of the nominal level for the asymptotic tests and bootstrap tests. For the distance d_∞ we also provide a comparison with the approach suggested by Gsteiger et al. (2011). In order to describe the method proposed by these authors note that it follows from (3.5) - (3.7) and an application of the Delta method [see Van der Vaart (1998)] that the variance of the prediction $m_1(x,\hat\beta_1) - m_2(x,\hat\beta_2)$ for difference of the two regression models at the point x is approximately normally distributed, that is

$$\frac{m_1(x,\hat\beta_1) - m_2(x,\hat\beta_2)}{\hat\tau_{n_1,n_2}(x,\hat\beta_1,\hat\beta_2)} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where

$$\hat\tau^2_{n_1,n_2}(x,\hat\beta_1,\hat\beta_2) = \sum_{\ell=1}^{2} \tfrac{1}{n_\ell}\big(\tfrac{\partial}{\partial\beta_\ell} m_\ell(x,\ \beta_\ell)\big|_{\beta_\ell=\hat\beta_\ell}\big)^T \hat\Sigma_\ell^{-1} \tfrac{\partial}{\partial\beta_\ell} m_\ell(x,\beta_\ell)\big|_{\beta_\ell=\hat\beta_\ell} \tag{5.1}$$

and $\hat\Sigma_\ell$ denotes the estimator of the variance in (3.4), which is obtained by replacing the parameters β_ℓ and $\zeta_{\ell,i}$ by their estimates $\hat\beta_\ell$, and $n_{\ell,i}/n_\ell$ ($\ell = 1,2$). Gsteiger et al. (2011) define a confidence band by

$$m_1(x,\hat\beta_1) - m_2(x,\hat\beta_2) \pm\ z_{1-\alpha}\hat\tau_{n_1,n_2}(x,\hat\beta_1,\hat\beta_2),$$

where $z_{1-\alpha}$ is the $(1-\alpha)$-quantile of the distribution of the random variable

$$D := \max_{x\in\mathcal{X}} \frac{\big|\frac{1}{\sqrt{n_1}}\frac{\partial}{\partial\beta_1}\big(m_1(x,\beta_1)\big|_{\beta_1=\hat\beta_1}\big)^T \hat\Sigma_1^{-1/2} Z_1 - \frac{1}{\sqrt{n_2}}\big(\frac{\partial}{\partial\beta_2} m_2(x,\beta_2)\big|_{\beta_2=\hat\beta_2}\big)^T \hat\Sigma_2^{-1/2} Z_2\big|}{\tau_{n_1,n_2}(x,\hat\beta_1,\hat\beta_2)}$$

and Z_1 and Z_2 are independent p_1- and p_2-dimensional standard normal distributed random variables, respectively. They proposed to determine this quantile by simulation, but they did not prove that this parametric bootstrap method is in fact a valid procedure. On the other hand, they demonstrated by means of a simulation study that the confidence bands obtained by this method have rather accurate coverage probabilities. A test for the hypotheses (4.6) is finally obtained by rejecting the null hypothesis, if the maximum (minimum) of the upper (lower) confidence band is smaller (larger) than ε ($-\varepsilon$). A particular advantage of this test is that it directly refers to the distance (2.4), which has a nice interpretation in applications. Moreover, in linear models (with normally distributed errors) this test keeps its level not only asymptotically but also for fixed sample size. However, the resulting test is extremely conservative and - as it will be demonstrated in Section 5.2 - has very low power compared to the methods proposed in this paper.

5.1 Tests based on the distance d_2

All presented results in this and the following section are based on 1000 simulation runs and the quantiles of the bootstrap tests have been obtained by $B = 300$ bootstrap replications. For the sake of brevity we restrict ourselves to a comparison of two shifted EMAX-models, i.e.

$$m_1(x, \alpha) = \delta + \frac{5x}{1+x}, \qquad m_2(x, \beta) = \frac{5x}{1+x}. \tag{5.2}$$

The dose range is given by the interval $\mathcal{X} = [0, 4]$ and an equal number of observations was taken at five dose levels $x_{\ell,1} = 0, x_{\ell,2} = 1$, $x_{\ell,3} = 2$, $x_{\ell,4} = 3$, $x_{\ell,5} = 4$ in both groups (that is $k_1 = k_2 = 5$). In Table 1 and 2 we present the simulated type I error of the bootstrap test (3.16) and the asymptotic test (3.12) respectively, where the threshold ε in the hypothesis (3.1) was chosen as $\varepsilon = 1$. Various configurations of σ_1^2, σ_2^2, n_1, n_2 and δ were considered. In the interior of the null hypothesis (that is $d_2 > \varepsilon$) the type one errors of the tests (3.12) and (3.16) are substantially smaller than the nominal level as predicted by Remark 3.2. For both tests we observe a rather precise approximation of the nominal level (even for small sample sizes) at the boundary of the null hypothesis (i.e. $\varepsilon = 1$). In some cases the asymptotic test (3.12) does not keep its 10%-level and for this reason we recommend to use the bootstrap test (3.16) to establish equivalence of two regression models with respect to the L^2-distance.

			$\alpha = 0.05$				$\alpha = 0.1$			
			(σ_1^2, σ_2^2)				(σ_1^2, σ_2^2)			
(n_1, n_2)	δ	d_2	$(0.25, 0.25)$	$(0.5, 0.5)$	$(0.25, 0.5)$	$(0.5, 1)$	$(0.25, 0.25)$	$(0.5, 0.5)$	$(0.25, 0.5)$	$(0.5, 1)$
(10, 10)	1	4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(10, 10)	0.75	2.25	0.004	0.002	0.001	0.002	0.000	0.002	0.000	0.012
(10, 10)	0.5	1	0.051	0.064	0.052	0.043	0.101	0.120	0.118	0.115
(10, 20)	1	4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(10, 20)	0.75	2.25	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005
(10, 20)	0.5	1	0.055	0.060	0.051	0.051	0.104	0.111	0.101	0.102
(20, 20)	1	4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(20, 20)	0.75	2.25	0.001	0.002	0.000	0.017	0.004	0.005	0.001	0.031
(20, 20)	0.5	1	0.057	0.058	0.050	0.067	0.125	0.107	0.097	0.127
(50, 50)	1	4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(50, 50)	0.75	2.25	0.001	0.000	0.000	0.000	0.002	0.000	0.000	0.000
(50, 50)	0.5	1	0.057	0.048	0.054	0.052	0.097	0.114	0.093	0.114

Table 1: *Simulated level of the d_2-bootstrap test* (3.16) *for the equivalence of two shifted EMAX models defined in* (5.2). *The threshold in* (3.1) *is chosen as* $\varepsilon = 1$.

			$\alpha = 0.05$				$\alpha = 0.1$			
			(σ_1^2, σ_2^2)				(σ_1^2, σ_2^2)			
(n_1, n_2)	δ	d_2	(0.25, 0.25)	(0.5, 0.5)	(0.25, 0.5)	(0.5, 1)	(0.25, 0.25)	(0.5, 0.5)	(0.25, 0.5)	(0.5, 1)
(10, 10)	1	4	0.002	0.002	0.002	0.000	0.000	0.002	0.003	0.005
(10, 10)	0.75	2.25	0.005	0.005	0.009	0.008	0.007	0.011	0.016	0.013
(10, 10)	0.5	1	0.080	0.042	0.049	0.032	0.102	0.061	0.071	0.050
(10, 20)	1	4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(10, 20)	0.75	2.25	0.007	0.012	0.007	0.015	0.017	0.015	0.012	0.017
(10, 20)	0.5	1	0.055	0.063	0.060	0.048	0.081	0.078	0.084	0.070
(20, 20)	1	4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(20, 20)	0.75	2.25	0.000	0.001	0.002	0.012	0.017	0.003	0.006	0.013
(20, 20)	0.5	1	0.060	0.066	0.080	0.066	0.090	0.091	0.096	0.092
(50, 50)	1	4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(50, 50)	0.75	2.25	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002
(50, 50)	0.5	1	0.041	0.058	0.052	0.086	0.071	0.087	0.073	0.117

Table 2: *Simulated level of the asymptotic d_2-test* (3.12) *for the equivalence of two shifted EMAX models defined in* (5.2). *The threshold in* (3.1) *is chosen as* $\varepsilon = 1$.

In Tables 3 and 4 we display the power of the two tests for various alternatives specified by the value δ in model (5.2). We observe a reasonable power of both tests in all cases under consideration. In the cases where the asymptotic test (3.12) keeps (or exceeds) its nominal level it is slightly more powerful than the bootstrap test (3.16). On the other hand the opposite behaviour can be observed in the cases where the asymptotic test is conservative (for example, if $\alpha = 10\%$, $n_1 = n_2 = 10$). We also note that the power of both tests is a decreasing function of the distance d_2, as predicted by the asymptotic theory.

			$\alpha = 0.05$				$\alpha = 0.1$			
			(σ_1^2, σ_2^2)				(σ_1^2, σ_2^2)			
(n_1, n_2)	δ	d_2	(0.25, 0.25)	(0.5, 0.5)	(0.25, 0.5)	(0.5, 1)	(0.25, 0.25)	(0.5, 0.5)	(0.25, 0.5)	(0.5, 1)
(10, 10)	0.25	0.25	0.210	0.118	0.134	0.080	0.300	0.212	0.256	0.214
(10, 10)	0.1	0.04	0.294	0.132	0.186	0.086	0.427	0.250	0.312	0.164
(10, 10)	0	0	0.351	0.145	0.176	0.090	0.467	0.286	0.340	0.160
(10, 20)	0.25	0.25	0.257	0.125	0.191	0.105	0.392	0.234	0.305	0.226
(10, 20)	0.1	0.04	0.395	0.164	0.254	0.121	0.535	0.305	0.395	0.238
(10, 20)	0	0	0.437	0.158	0.291	0.148	0.598	0.290	0.474	0.260
(20, 20)	0.25	0.25	0.392	0.171	0.225	0.140	0.534	0.302	0.382	0.230
(20, 20)	0.1	0.04	0.560	0.308	0.418	0.165	0.720	0.460	0.562	0.287
(20, 20)	0	0	0.610	0.314	0.390	0.180	0.757	0.462	0.555	0.307
(50, 50)	0.25	0.25	0.724	0.460	0.554	0.245	0.825	0.595	0.825	0.452
(50, 50)	0.1	0.04	0.961	0.691	0.821	0.485	0.982	0.824	0.973	0.647
(50, 50)	0	0	0.984	0.734	0.865	0.508	0.998	0.861	0.999	0.684

Table 3: *Simulated power of the d_2-bootstrap test* (3.16) *for the equivalence of two shifted EMAX models defined in* (5.2). *The threshold in* (3.1) *is chosen as* $\varepsilon = 1$.

			$\alpha = 0.05$				$\alpha = 0.1$			
			(σ_1^2, σ_2^2)				(σ_1^2, σ_2^2)			
(n_1, n_2)	δ	d_2	(0.25, 0.25)	(0.5, 0.5)	(0.25, 0.5)	(0.5, 1)	(0.25, 0.25)	(0.5, 0.5)	(0.25, 0.5)	(0.5, 1)
(10, 10)	0.25	0.25	0.264	0.103	0.175	0.082	0.311	0.131	0.217	0.102
(10, 10)	0.1	0.04	0.351	0.139	0.196	0.068	0.431	0.183	0.247	0.100
(10, 10)	0	0	0.381	0.120	0.222	0.078	0.468	0.168	0.279	0.097
(10, 20)	0.25	0.25	0.305	0.147	0.256	0.112	0.382	0.192	0.317	0.153
(10, 20)	0.1	0.04	0.468	0.218	0.359	0.135	0.536	0.268	0.438	0.170
(10, 20)	0	0	0.510	0.220	0.358	0.110	0.570	0.272	0.455	0.161
(20, 20)	0.25	0.25	0.423	0.271	0.321	0.172	0.493	0.328	0.341	0.216
(20, 20)	0.1	0.04	0.640	0.328	0.501	0.208	0.716	0.407	0.585	0.260
(20, 20)	0	0	0.690	0.351	0.501	0.206	0.781	0.438	0.573	0.272
(50, 50)	0.25	0.25	0.659	0.475	0.534	0.382	0.740	0.562	0.649	0.446
(50, 50)	0.1	0.04	0.965	0.750	0.868	0.556	0.974	0.813	0.911	0.637
(50, 50)	0	0	0.980	0.848	0.937	0.601	0.991	0.893	0.946	0.668

Table 4: *Simulated power of the asymptotic d_2- test* (3.12) *for the equivalence of two shifted EMAX models defined in* (5.2)*. The threshold in* (3.1) *is chosen as $\varepsilon = 1$.*

5.2 Tests based on the distance d_∞

In this section we investigate the maximum deviation distance and also provide a comparison with the test for the hypotheses (4.6), which has recently been proposed by Gsteiger et al. (2011). We begin with a comparison of an EMAX and an exponential model, that is

$$m_1(x, \beta_1) = 1 + \frac{2x}{1+x}\ , \qquad m_2(x, \beta_2) = \delta + 2.2 \cdot (\exp\left(\tfrac{x}{8}\right) - 1). \tag{5.3}$$

The dose range is given by the interval $\mathcal{X} = [0, 4]$ and an equal number of patients was allocated at five dose levels $x_{\ell,1} = 0, x_{\ell,2} = 1,\ x_{\ell,3} = 2,\ x_{\ell,4} = 3,\ x_{\ell,5} = 4$ in both groups (that is $k_1 = k_2 = 5$). In Table 5 we display the simulated rejection probabilities of the the bootstrap test (4.8) under the null hypothesis in (4.6), where $\varepsilon = 1$. The results for the asymptotic test are given in Table 6. We note that this test can be used in the present context, because in example (5.3) the cardinality of the set $\mathcal{E}$ of extremal points of the "true" difference $m_2(x, \beta_1) - m_2(x, \beta_2)$ is one. Thus, if the unique extremal point has been estimated, we obtain by (4.7) an estimate, say $\hat{\sigma}_\infty^2$, of the asymptotic variance of the statistic $\hat{d}_\infty$. The null hypothesis is now rejected (at asymptotic level α), whenever

$$\hat{d}_\infty < \varepsilon + \frac{\hat{\sigma}_\infty}{\sqrt{n}} u_\alpha\ , \tag{5.4}$$

where u_α is $\alpha-$quantile of the standard normal distribution. We observe that the bootstrap test (4.8) keeps it nominal level at the boundary of the null hypothesis, where the level is smaller in the interior (this confirms the theoretical results from Section 4). The approximation is less precise for small sample sizes. Compared to the d_2-bootstrap test, the test (4.8) is conservative. On the other hand the asymptotic test (5.4) is extremely conservative, even for relative large sample sizes (see Table 6). In Table 5 we also display the rejecting probabilities of the corresponding test of Gsteiger et al. (2011) in brackets. This test is described in Section 2 and we observe that it is extremely conservative. The level of the test of Gsteiger et al. (2011)

is practically 0 in nearly all cases under consideration.

The simulated power of the bootstrap and the asymptotic d_∞-test are displayed in Table 7 and 8. We observe a substantially better performance of the bootstrap test (4.8) in all cases of consideration. In Table 7 we also display the rejecting probabilities of the test of Gsteiger et al. (2011) in brackets. This test has practically no power, and the method proposed in this paper yields a substantial improvement.

			$\alpha = 0.05$				$\alpha = 0.1$			
			(σ_1^2, σ_2^2)				(σ_1^2, σ_2^2)			
(n_1, n_2)	δ	d_∞	(0.25, 0.25)	(0.5, 0.5)	(0.25, 0.5)	(0.5, 1)	(0.25, 0.25)	(0.5, 0.5)	(0.25, 0.5)	(0.5, 1)
(10, 10)	0.25	1.5	0.000 (0.000)	0.001 (0.000)	0.000 (0.000)	0.001 (0.000)	0.000 (0.000)	0.004 (0.000)	0.000 (0.000)	0.007 (0.000)
(10, 10)	0.5	1.25	0.000 (0.000)	0.003 (0.000)	0.001 (0.000)	0.010 (0.000)	0.003 (0.000)	0.014 (0.000)	0.005 (0.000)	0.018 (0.000)
(10, 10)	0.75	1	0.035 (0.001)	0.027 (0.000)	0.023 (0.000)	0.032 (0.000)	0.075 (0.000)	0.095 (0.000)	0.066 (0.001)	0.072 (0.000)
(10, 20)	0.25	1.5	0.000 (0.000)	0.002 (0.000)	0.000 (0.000)	0.001 (0.000)	0.000 (0.000)	0.002 (0.000)	0.000 (0.000)	0.002 (0.000)
(10, 20)	0.5	1.25	0.004 (0.000)	0.003 (0.000)	0.001 (0.000)	0.008 (0.000)	0.008 (0.000)	0.009 (0.000)	0.009 (0.000)	0.020 (0.000)
(10, 20)	0.75	1	0.045 (0.000)	0.026 (0.000)	0.028 (0.000)	0.031 (0.000)	0.112 (0.002)	0.086 (0.000)	0.079 (0.000)	0.080 (0.001)
(20, 20)	0.25	1.5	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
(20, 20)	0.5	1.25	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)	0.002 (0.000)	0.003 (0.000)	0.007 (0.000)	0.001 (0.000)	0.007 (0.000)
(20, 20)	0.75	1	0.034 (0.000)	0.021 (0.000)	0.021 (0.000)	0.018 (0.000)	0.082 (0.000)	0.074 (0.000)	0.063 (0.000)	0.060 (0.000)
(50, 50)	0.25	1.5	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
(50, 50)	0.5	1.25	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.001 (0.000)	0.000 (0.000)	0.000 (0.000)
(50, 50)	0.75	1	0.051 (0.000)	0.026 (0.000)	0.022 (0.000)	0.015 (0.000)	0.106 (0.000)	0.078 (0.000)	0.062 (0.000)	0.057 (0.000)

Table 5: *Simulated level of the d_∞-bootstrap test* (4.8) *for the equivalence of an EMAX and an exponential model defined by* (5.3). *The threshold in* (4.6) *is chosen as* $\varepsilon = 1$.

			$\alpha = 0.05$				$\alpha = 0.1$			
			(σ_1^2, σ_2^2)				(σ_1^2, σ_2^2)			
(n_1, n_2)	δ	d_∞	(0.25, 0.25)	(0.5, 0.5)	(0.25, 0.5)	(0.5, 1)	(0.25, 0.25)	(0.5, 0.5)	(0.25, 0.5)	(0.5, 1)
(10, 10)	0.25	1.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(10, 10)	0.5	1.25	0.001	0.001	0.000	0.000	0.003	0.004	0.001	0.000
(10, 10)	0.75	1	0.012	0.005	0.003	0.000	0.029	0.010	0.001	0.001
(10, 20)	0.25	1.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
(10, 20)	0.5	1.25	0.000	0.005	0.001	0.000	0.000	0.007	0.003	0.001
(10, 20)	0.75	1	0.019	0.006	0.009	0.004	0.038	0.014	0.023	0.009
(20, 20)	0.25	1.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(20, 20)	0.5	1.25	0.000	0.000	0.001	0.002	0.000	0.001	0.001	0.001
(20, 20)	0.75	1	0.011	0.036	0.009	0.004	0.033	0.025	0.027	0.0016
(50, 50)	0.25	1.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(50, 50)	0.5	1.25	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.001
(50, 50)	0.75	1	0.016	0.015	0.012	0.008	0.039	0.039	0.041	0.026

Table 6: *Simulated level of the asymptotic d_∞-test* (5.4) *for the equivalence of an EMAX and an exponential model defined by* (5.3). *The threshold in* (4.6) *is chosen as* $\varepsilon = 1$.

			$\alpha = 0.05$				$\alpha = 0.1$			
			(σ_1^2, σ_2^2)				(σ_1^2, σ_2^2)			
(n_1, n_2)	δ	d_∞	(0.25, 0.25)	(0.5, 0.5)	(0.25, 0.5)	(0.5, 1)	(0.25, 0.25)	(0.5, 0.5)	(0.25, 0.5)	(0.5, 1)
(10, 10)	1	0.75	0.160 (0.002)	0.093 (0.000)	0.125 (0.000)	0.102 (0.000)	0.297 (0.005)	0.225 (0.000)	0.229 (0.001)	0.216 (0.000)
(10, 10)	1.5	0.5	0.237 (0.003)	0.133 (0.000)	0.164 (0.000)	0.120 (0.000)	0.383 (0.008)	0.231 (0.000)	0.309 (0.005)	0.232 (0.000)
(10, 20)	1	0.75	0.185 (0.003)	0.123 (0.000)	0.159 (0.000)	0.120 (0.000)	0.320 (0.006)	0.226 (0.000)	0.283 (0.003)	0.219 (0.000)
(10, 20)	1.5	0.5	0.300 (0.006)	0.175 (0.000)	0.269 (0.000)	0.133 (0.000)	0.457 (0.007)	0.305 (0.000)	0.414 (0.005)	0.255 (0.000)
(20, 20)	1	0.75	0.214 (0.010)	0.138 (0.000)	0.171 (0.002)	0.118 (0.000)	0.393 (0.018)	0.271 (0.001)	0.345 (0.002)	0.232 (0.001)
(20, 20)	1.5	0.5	0.401 (0.020)	0.229 (0.000)	0.363 (0.003)	0.168 (0.000)	0.604 (0.047)	0.398 (0.002)	0.523 (0.010)	0.317 (0.000)
(50, 50)	1	0.75	0.473 (0.122)	0.255 (0.013)	0.363 (0.043)	0.200 (0.001)	0.662 (0.178)	0.436 (0.037)	0.562 (0.072)	0.357 (0.003)
(50, 50)	1.5	0.5	0.851 (0.262)	0.550 (0.022)	0.740 (0.061)	0.385 (0.002)	0.927 (0.361)	0.716 (0.068)	0.845 (0.148)	0.565 (0.013)

Table 7: *Simulated power of the* d_∞*-bootstrap test* (4.8) *for the equivalence of an EMAX and an exponential model defined by* (5.3). *The threshold in* (4.6) *is chosen as* $\varepsilon = 1$.

			$\alpha = 0.05$				$\alpha = 0.1$			
			(σ_1^2, σ_2^2)				(σ_1^2, σ_2^2)			
(n_1, n_2)	δ	d_∞	(0.25, 0.25)	(0.5, 0.5)	(0.25, 0.5)	(0.5, 1)	(0.25, 0.25)	(0.5, 0.5)	(0.25, 0.5)	(0.5, 1)
(10, 10)	1	0.75	0.042	0.006	0.011	0.001	0.109	0.017	0.046	0.007
(10, 10)	1.5	0.5	0.064	0.008	0.014	0.002	0.140	0.026	0.047	0.009
(10, 20)	1	0.75	0.114	0.014	0.048	0.004	0.199	0.047	0.106	0.023
(10, 20)	1.5	0.5	0.129	0.018	0.059	0.006	0.228	0.052	0.127	0.025
(20, 20)	1	0.75	0.151	0.036	0.064	0.009	0.285	0.093	0.170	0.035
(20, 20)	1.5	0.5	0.209	0.060	0.104	0.012	0.360	0.120	0.202	0.044
(50, 50)	1	0.75	0.417	0.206	0.303	0.111	0.569	0.337	0.440	0.221
(50, 50)	1.5	0.5	0.706	0.267	0.408	0.146	0.826	0.462	0.630	0.303

Table 8: *Simulated power of the asymptotic* d_∞*-test* (5.4) *for the equivalence of an EMAX and an exponential model defined by* (5.3). *The threshold in* (4.6) *is chosen as* $\varepsilon = 1$.

We conclude this section with an investigation of the models in (5.2). In this case the set of extremal points of the "true" difference is given by $\mathcal{E} = [0, 4]$ and an asymptotic test based on the maximum deviation is not available. In Table 9 we display the rejection probabilities of the bootstrap test (4.8) under the null hypothesis. The numbers in brackets show again the corresponding values for the test Gsteiger et al. (2011). We observe that in this case both tests are conservative.
Corresponding results under the alternative are shown in Table 10, where it is demonstrated that the bootstrap test (4.8) yields again a substantial improvement in power compared to the test of Gsteiger et al. (2011). While this test has practically no power, the new bootstrap test proposed in this paper is able to establish equivalence between the curves with a reasonable type II error, if the total sample size is larger than 50.

		$\alpha = 0.05$				$\alpha = 0.1$			
		(σ_1^2, σ_2^2)				(σ_1^2, σ_2^2)			
(n_1, n_2)	$\delta = d_\infty$	$(0.25, 0.25)$	$(0.5, 0.5)$	$(0.25, 0.5)$	$(0.5, 1)$	$(0.25, 0.25)$	$(0.5, 0.5)$	$(0.25, 0.5)$	$(0.5, 1)$
(10, 10)	1	0.000 (0.000)	0.004 (0.000)	0.001 (0.000)	0.005 (0.000)	0.007 (0.000)	0.019 (0.000)	0.010 (0.000)	0.024 (0.000)
(10, 10)	0.75	0.000 (0.000)	0.008 (0.000)	0.006 (0.000)	0.015 (0.000)	0.013 (0.002)	0.041 (0.000)	0.020 (0.000)	0.043 (0.000)
(10, 10)	0.5	0.015 (0.001)	0.040 (0.000)	0.016 (0.000)	0.030 (0.000)	0.050 (0.005)	0.104 (0.000)	0.054 (0.002)	0.074 (0.001)
(10, 20)	1	0. 000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.003 (0.000)	0.005 (0.000)	0.002 (0.000)	0.003 (0.000)	0.012 (0.000)
(10, 20)	0.75	0.001 (0.000)	0.004 (0.000)	0.000 (0.000)	0.006 (0.000)	0.005 (0.000)	0.023 (0.000)	0.006 (0.000)	0.036 (0.000)
(10, 20)	0.5	0.018 (0.000)	0.016 (0.000)	0.012 (0.000)	0.030 (0.000)	0.045 (0.000)	0.051 (0.000)	0.037 (0.000)	0.077 (0.000)
(20, 20)	1	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.001 (0.000)	0.000 (0.000)	0.004 (0.000)	0.006 (0.000)	0.006 (0.000)
(20, 20)	0.75	0.000 (0.000)	0.002 (0.000)	0.000 (0.000)	0.005 (0.000)	0.003 (0.000)	0.010 (0.002)	0.002 (0.000)	0.016 (0.000)
(20, 20)	0.5	0.006 (0.001)	0.019 (0.000)	0.016 (0.000)	0.026 (0.000)	0.027 (0.001)	0.051 (0.000)	0.046 (0.000)	0.067 (0.000)
(50, 50)	1	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.001 (0.000)	0.000 (0.000)
(50, 50)	0.75	0.006 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.004 (0.000)	0.007 (0.000)	0.002 (0.000)	0.006 (0.000)
(50, 50)	0.5	0.003 (0.000)	0.005 (0.000)	0.004 (0.000)	0.008 (0.000)	0.018 (0.000)	0.027 (0.000)	0.034 (0.000)	0.031 (0.000)

Table 9: *Simulated level of the bootstrap* d_∞*-test* (4.8) *for the equivalence of two shifted EMAX-models defined by* (5.2)*. The threshold in* (4.6) *is chosen as* $\varepsilon = 0.5$*.*

		$\alpha = 0.05$				$\alpha = 0.1$			
		(σ_1^2, σ_2^2)				(σ_1^2, σ_2^2)			
(n_1, n_2)	$\delta = d_\infty$	$(0.25, 0.25)$	$(0.5, 0.5)$	$(0.25, 0.5)$	$(0.5, 1)$	$(0.25, 0.25)$	$(0.5, 0.5)$	$(0.25, 0.5)$	$(0.5, 1)$
(10, 10)	0.25	0.062 (0.000)	0.050 (0.000)	0.053 (0.000)	0.059 (0.000)	0.147 (0.000)	0.118 (0.000)	0.118 (0.000)	0.129
(10, 10)	0.1	0.100 (0.000)	0.070 (0.000)	0.099 (0.000)	0.066 (0.000)	0.195 (0.000)	0.137 (0.000)	0.190 (0.000)	0.129 (0.000)
(10, 10)	0	0.109 (0.000)	0.090 (0.000)	0.092 (0.000)	0.067 (0.000)	0.216 (0.000)	0.143 (0.000)	0.176 (0.000)	0.139(0.000)
(10, 20)	0.25	0.077 (0.000)	0.077 (0.000)	0.074 (0.000)	0.062 (0.000)	0.157 (0.000)	0.142 (0.000)	0.141 (0.000)	0.130 (0.000)
(10, 20)	0.1	0.118 (0.001)	0.077 (0.001)	0.100 (0.000)	0.078 (0.000)	0.227 (0.002)	0.163 (0.002)	0.176 (0.000)	0.148 (0.000)
(10, 20)	0	0.151 (0.001)	0.078 (0.001)	0.118 (0.000)	0.077 (0.000)	0.275 (0.004)	0.165 (0.003)	0.213 (0.000)	0.151 (0.000)
(20, 20)	0.25	0.085 (0.000)	0.060 (0.000)	0.076 (0.000)	0.061 (0.000)	0.171 (0.001)	0.134 (0.001)	0.162 (0.000)	0.121 (0.000)
(20, 20)	0.1	0.158 (0.000)	0.090 (0.000)	0.112 (0.000)	0.079 (0.000)	0.309 (0.003)	0.184 (0.002)	0.220 (0.001)	0.174 (0.000)
(20, 20)	0	0.178 (0.003)	0.108 (0.001)	0.120 (0.003)	0.083 (0.000)	0.324 (0.012)	0.209 (0.001)	0.219 (0.008)	0.157 (0.000)
(50, 50)	0.25	0.162 (0.000)	0.086 (0.000)	0.098 (0.000)	0.063 (0.000)	0.283 (0.000)	0.178 (0.000)	0.218 (0.000)	0.153 (0.000)
(50, 50)	0.1	0.390 (0.019)	0.212 (0.002)	0.232 (0.012)	0.137 (0.001)	0.568 (0.054)	0.349 (0.006)	0.398 (0.025)	0.265 (0.003)
(50, 50)	0	0.457 (0.096)	0.211 (0.012)	0.266 (0.032)	0.151 (0.001)	0.630 (0.194)	0.363 (0.033)	0.438 (0.069)	0.261 (0.004)

Table 10: *Simulated power of the bootstrap* d_∞*-test* (4.8) *for the equivalence of two shifted EMAX-models defined by* (5.2)*. The threshold in* (4.6) *is chosen as* $\varepsilon = 0.5$*.*

Acknowledgements This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Project C1) of the German Research Foundation (DFG). Kathrin Möllenhoff's research has received funding from the European Union Seventh Framework Programme [FP7 2007–2013] under grant agreement no 602552 (IDEAL - Integrated Design and Analysis of small population group trials). The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise.

References

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

Chow, S.-C. and Liu, P.-J. (1992). *Design and Analysis of Bioavailability and Bioequivalence Studies.* Marcel Dekker, New York.

Gsteiger, S., Bretz, F., and 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., and Pigeot, I. (2007). *Bioequivalence Studies in Drug Development Methods and Applications. Statistics in Practice*. John Wiley & Sons, New York.

Kosorok, M. R. (2007). *Introduction to empirical processes and semiparametric inference*. Springer Science & Business Media.

Liu, W., Bretz, F., Hayter, A. J., and 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.

Raghavachari, M. (1973). Limiting distributions of Kolmogorov-Smirnov type statistics under the alternative. *Annals of Statistics*, 1(1):67–73.

Van der Vaart, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

6 Appendix: Technical details

The theoretical results of this paper are proved under the following assumptions.

Assumption 6.1. *The errors $\varepsilon_{i,j,\ell}$ have finite variance σ_ℓ^2.*

Assumption 6.2. *The covariate region $\mathcal{X} \subset \mathbb{R}^d$ is compact and the number and location of dose levels k_ℓ does not depend on n_ℓ, $\ell = 1, 2$.*

Assumption 6.3. *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 6.4. *The regression functions m_1 and m_2 are twice continuously differentiable with respect to the parameters for all b_1, b_2 in neighbourhoods of the true parameters β_1, β_2 and all $x \in \mathcal{X}$. The functions $(x, b_\ell) \mapsto m_\ell(x, b_\ell)$ and their first two derivatives are continuous on $\mathcal{X} \times B_\ell$.*

Assumption 6.5. *Defining*

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

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

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

In particular, under Assumptions 6.1 - 6.5 the least squares estimator can be linearized. To be precise, consider arbitrary sequences $(\beta_{\ell,n})_{n\in\mathbb{N}}$ and $(\sigma_{\ell,n})_{n\in\mathbb{N}}$ in B_ℓ and $\mathbb{R}^+$ such that $\beta_{\ell,n} \to \beta_\ell$ and $\sigma_{\ell,n} \to \sigma_\ell > 0$ as $n_1, n_2 \to \infty$ ($\ell = 1, 2$) and denote by $Y_{\ell,i,j}^{(n)}$ data of the form given in (2.1), (2.2) with β_ℓ replaced by $\beta_{\ell,n}$ and $\varepsilon_{\ell,i,j}$ independent and identically distributed (for each fixed n) with mean 0 and finite variances $\sigma_{\ell,n}^2$. Then the least squares estimators $\hat{\beta}_\ell^{(n)}$ computed from $Y_{\ell,i,j}^{(n)}$ satisfy

$$\sqrt{n_\ell}\, (\hat{\beta}_\ell - \beta_{\ell,n}) = \frac{1}{\sqrt{n_\ell}} \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} \phi_{\ell,i,j} + o_{\mathbb{P}}(1), \qquad \ell = 1, 2, \tag{6.1}$$

where the functions $\phi_{\ell,i,j}^{(n)}$ are given by

$$\phi_{\ell,i,j} = \tfrac{\varepsilon_{\ell,i,j}}{\sigma_{\ell,n}^2} \Sigma_{\ell,n}^{-1} \tfrac{\partial}{\partial b_l} m_\ell(x_{\ell,i}, b_\ell)\big|_{b_l=\beta_{\ell,n}}, \qquad \ell = 1, 2, \tag{6.2}$$

and $\Sigma_{\ell,n}$ takes the form

$$\Sigma_{\ell,n} = \frac{1}{\sigma_{\ell,n}^2} \sum_{i=1}^{k_\ell} \zeta_{\ell,i} \tfrac{\partial}{\partial b_l} m_\ell(x_{\ell,i}, b_\ell)\big|_{b_l=\beta_{\ell,n}} \big(\tfrac{\partial}{\partial b_l} m_\ell(x_{\ell,i}, b_\ell)\big|_{b_l=\beta_{\ell,n}}\big)^T, \quad \ell = 1, 2. \tag{6.3}$$

6.1 Proof of Theorem 3.1:

Let $\ell_\infty(\mathcal{X})$ denote the space of all bounded real valued functions of the form $g : \mathcal{X} \to \mathbb{R}$. The mapping

$$\Phi : \begin{cases} \mathbb{R}^{p_1+p_2} \to \ell_\infty(\mathcal{X}) \\ (\theta_1, \theta_2) \mapsto \Phi(\theta_1, \theta_2) : \begin{cases} \mathcal{X} \mapsto \mathbb{R} \\ x \mapsto \big(\frac{\partial}{\partial \beta_1} m_1(x, \beta_1)\big)^T \theta_1 - \big(\frac{\partial}{\partial \beta_2} m_1(x, \beta_2)\big)^T \theta_2 \end{cases} \end{cases}, \tag{6.4}$$

is continuous due to Assumptions 6.2-6.4, where we use the Euclidean and the supremum norm on $\mathbb{R}^{p_1+p_2}$ and $\ell_\infty(\mathcal{X})$, respectively. Consequently, the continuous mapping theorem [see Van der Vaart (1998)] and (3.7) yield that the process

$$\{\sqrt{n} G_n(x)\}_{x \in \mathcal{X}} := \big\{\sqrt{n}\big((\tfrac{\partial}{\partial \beta_1} m_1(x, \beta_1))^T (\hat{\beta}_1 - \beta_1) - (\tfrac{\partial}{\partial \beta_2} m_2(x, \beta_2))^T (\hat{\beta}_2 - \beta_2)\big)\big\}_{x \in \mathcal{X}}$$

converges weakly to a centered Gaussian process $\{G(x)\}_{x \in \mathcal{X}}$ in $\ell_\infty(\mathcal{X})$, which is defined by

$$G(x) = \big(\tfrac{\partial}{\partial \beta_1} m_1(x, \beta_1)\big)^T \sqrt{\lambda} \Sigma_1^{-1/2} Z_1 - \big(\tfrac{\partial}{\partial \beta_2} m_2(x, \beta_2)\big)^T \sqrt{\tfrac{\lambda}{\lambda - 1}} \Sigma_2^{-1/2} Z_2, \tag{6.5}$$

where Z_1 and Z_2 are p_1- and p_2-dimensional standard normal distributed random variables, respectively, i.e. $Z_\ell \sim \mathcal{N}(0, I_{p_\ell})$, $\ell = 1, 2$. A straightforward calculation shows that the covariance kernel of the process $\{G(x)\}_{x \in \mathcal{X}}$ is given by (3.11). Now a Taylor expansion gives

$$p_n(x) := \big(m_1(x, \hat{\beta}_1) - m_1(x, \beta_1)\big) - \big(m_2(x, \hat{\beta}_2) - m_2(x, \beta_2)\big) = G_n(x) + o_{\mathbb{P}}\Big(\sqrt{\frac{1}{n}}\Big), \tag{6.6}$$

uniformly with respect to $x \in \mathcal{X}$, and it therefore follows that

$$\{\sqrt{n} p_n(x)\}_{x \in \mathcal{X}} \xrightarrow{\mathcal{D}} \{G(x)\}_{x \in \mathcal{X}} . \tag{6.7}$$

Recalling the definition of $\Delta(x, \beta_1, \beta_2)$ in (2.6), observing the representation

$$\begin{aligned} \sqrt{n}(\hat{d}_2 - d_2) &= \sqrt{n}\Big(\int_{\mathcal{X}} \Delta^2(x, \hat{\beta}_1, \hat{\beta}_2) dx - \int_{\mathcal{X}} \Delta^2(x, \beta_1, \beta_2) dx\Big) \\ &= \sqrt{n} \int_{\mathcal{X}} \big(p_n(x) + 2\Delta(x, \beta_1, \beta_2)\big) p_n(x) dx \\ &= \int_{\mathcal{X}} \sqrt{n} p_n^2(x) dx + 2\sqrt{n} \int_{\mathcal{X}} \Delta(x, \beta_1, \beta_2) p_n(x) dx, \end{aligned}$$

and the continuous mapping theorem we therefore obtain

$$\sqrt{n}(\hat{d}_2 - d_2) \xrightarrow{\mathcal{D}} 2 \int_{\mathcal{X}} \Delta(x, \beta_1, \beta_2) G(x) dx,$$

where G denotes the Gaussian process defined in (6.5). Now it is easy to see that the distribution on the right-hand side is a centered normal distribution with variance $\sigma^2_{d_2}$ defined in (3.10). This completes the proof of Theorem 3.1. □

6.2 Proof of Theorem 3.4

Proof of (1). First we will derive the asymptotic distribution of the bootstrap estimators $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$. Then we use similar arguments as given in the proof of Theorem 3.1 to derive the asymptotic distribution of the statistic $\hat{d}_2^*$ (appropriately standardized). Finally, in a third step, we establish the statement (3.17).

Recall the definition of the estimators in (3.13) and note that it follows from Assumptions 6.1-6.5 that under the null hypothesis $H_0 : d_2 \geq \varepsilon$

$$\hat{\hat{\beta}}_\ell \xrightarrow{\mathbb{P}} \beta_\ell \quad \ell = 1, 2, \quad \text{whenever} \quad d_2 \geq \varepsilon. \tag{6.8}$$

For $\ell = 1, 2$ let $\chi_\ell = \sigma(Y_{\ell,i,j} | i = 1, \ldots, k_\ell, j = 1, \ldots, n_{\ell,i})$ denote the σ-field generated by the random variables $\{Y_{\ell,i,j} | i = 1, \ldots, k_\ell, j = 1, \ldots, n_{\ell,i}\}$ and $\chi := \sigma(\chi_1, \chi_2)$ (note that we do not display the dependence of these quantities on the sample size). Given (6.8) and the consistency of $\hat{\sigma}_\ell$, the discussion after Assumption 6.5 yields

$$\sqrt{n_\ell}(\hat{\beta}_\ell^* - \hat{\hat{\beta}}_\ell) = \frac{1}{\sqrt{n_\ell}} \sum_{i=1}^{k_\ell} \sum_{j=1}^{n_{\ell,i}} \phi_{\ell,i,j}^* \ + \ o_{\mathbb{P}}(1) \ , \quad \ell = 1, 2, \tag{6.9}$$

where the quantities $\phi_{\ell,i,j}^*$ are given by

$$\phi_{\ell,i,j}^* = \tfrac{\varepsilon_{\ell,i,j}^*}{\hat{\sigma}_\ell^2} \hat{\hat{\Sigma}}_\ell^{-1} \tfrac{\partial}{\partial \beta_\ell} m_\ell(x_{\ell,i,}, \beta_\ell) \big|_{\beta_\ell = \hat{\hat{\beta}}_\ell}, \tag{6.10}$$

and the $p_1 \times p_1$ and $p_2 \times p_2$ dimensional matrices $\hat{\hat{\Sigma}}_1^{-1}$ and $\hat{\hat{\Sigma}}_2^{-1}$ are defined by

$$\hat{\hat{\Sigma}}_\ell = \frac{1}{\hat{\sigma}_\ell^2} \sum_{i=1}^{k_\ell} \zeta_{\ell,i} \big(\tfrac{\partial}{\partial \beta_\ell} m_\ell(x_{\ell,i,}, \beta_\ell) \big|_{\beta_\ell = \hat{\hat{\beta}}_\ell} \big) \big(\tfrac{\partial}{\partial \beta_\ell} m_\ell(x_{\ell,i,}, \beta_\ell) \big|_{\beta_\ell = \hat{\hat{\beta}}_\ell} \big)^T.$$

This yields the representation

$$\sqrt{n_\ell}(\hat{\beta}_\ell^* - \hat{\hat{\beta}}_\ell) = \sum_{i=1}^{k_\ell} \frac{1}{\hat{\sigma}_\ell} \hat{\hat{\Sigma}}_\ell^{-1} \tfrac{\partial}{\partial \beta_\ell} m_\ell(x_{\ell,i,}, \beta_\ell) \big|_{\beta_\ell = \hat{\hat{\beta}}_\ell} \frac{1}{\sqrt{n_\ell}} \sum_{j=1}^{n_{\ell,i}} \tfrac{\varepsilon_{\ell,i,j}^*}{\hat{\sigma}_\ell} \ + \ o_{\mathbb{P}}(1) \ , \quad \ell = 1, 2.$$

Since by construction the $\frac{\varepsilon_{\ell,i,j}^*}{\hat{\sigma}_\ell}$ are i.i.d. with unit variance and independent of χ, the classical central limit theorem implies that, conditionally on χ in probability

$$\sqrt{n_\ell} \cdot \Sigma_\ell^{\frac{1}{2}} (\hat{\beta}_\ell^* - \hat{\hat{\beta}}_\ell) \xrightarrow{D} \mathcal{N}(0, I_{p_\ell}) \ , \qquad \ell = 1, 2, \tag{6.11}$$

where the matrix Σ_ℓ is defined in (3.4).

We will now use this result to derive the weak convergence of the statistic

$$\hat{d}_2^* = d_2(\hat{\beta}_1^*, \hat{\beta}_2^*) = \int_{\mathcal{X}} (m_1(x, \hat{\beta}_1^*) - m_2(x, \hat{\beta}_2^*))^2 dx.$$

For this purpose we can proceed as in the proof of Theorem 3.1, where we discussed the weak convergence of the statistic $\hat{d}_2$. Recall that under the null hypothesis condition (6.8) is satisfied. It then follows by a Taylor expansion that

$$p_n^*(x) := \big(m_1(x,\hat{\beta}_1^*) - m_1(x,\hat{\hat{\beta}}_1)\big) - \big(m_2(x,\hat{\beta}_2^*) - m_2(x,\hat{\hat{\beta}}_2)\big) = G_n^*(x) + o_{\mathbb{P}}\Big(\frac{1}{\sqrt{n}}\Big) \quad (6.12)$$

(uniformly with respect to $x \in \mathcal{X}$), where the process $G_n^*(x)$ is given by

$$G_n^*(x) = (\tfrac{\partial}{\partial \beta_1} m_1(x,\beta_1))^T(\hat{\beta}_1^* - \hat{\hat{\beta}}_1) - (\tfrac{\partial}{\partial \beta_2} m_2(x,\beta_2))^T(\hat{\beta}_2^* - \hat{\hat{\beta}}_2). \quad (6.13)$$

Now the same arguments as given in the proof of Theorem 3.1 together with (6.11) - (6.13) and Proposition 10.7 in Kosorok (2007) show that, conditionally on χ in probability

$$\big\{\sqrt{n} p_n^*(x)\big\}_{x\in\mathcal{X}} \xrightarrow{\mathcal{D}} \{G(x)\}_{x\in\mathcal{X}} \quad (6.14)$$

where $\{G(x)\}_{x\in\mathcal{X}}$ is the centered Gaussian process defined in (6.5). A further application of Proposition 10.7 in Kosorok (2007) therefore now yields

$$\frac{\sqrt{n}}{\sigma_{d_2}}(\hat{d}_2^* - \hat{\hat{d}}_2) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad (6.15)$$

conditionally on χ in probability.
Now recall that $\hat{q}_\alpha$ is the α-quantile of the bootstrap statistic $\hat{d}_2^*$ conditionally on χ and note that, almost surely,

$$\alpha = \mathbb{P}(\hat{d}_2^* < \hat{q}_\alpha |\, \chi) = \mathbb{P}\Big(\frac{\sqrt{n}(\hat{d}_2^* - \hat{\hat{d}}_2)}{\sigma_{d_2}} < \frac{\sqrt{n}(\hat{q}_\alpha - \hat{\hat{d}}_2)}{\sigma_{d_2}} \Big|\, \chi\Big). \quad (6.16)$$

Letting

$$\hat{p}_\alpha := \frac{\sqrt{n}(\hat{q}_\alpha - \hat{\hat{d}}_2)}{\sigma_{d_2}}$$

it follows from (6.15), (6.16) and Lemma 21.2 in Van der Vaart (1998) that

$$\hat{p}_\alpha \xrightarrow{\mathbb{P}} u_\alpha, \quad (6.17)$$

where u_α denotes the α-quantile of the standard normal distribution. This relation implies for any $\alpha < 0.5$ that

$$\lim_{n\to\infty} \mathbb{P}(\hat{q}_\alpha - \hat{\hat{d}}_2 > 0\ ) = \lim_{n\to\infty} \mathbb{P}\,(\hat{p}_\alpha > 0) = 0. \quad (6.18)$$

After these preparations we are able to prove the first part of Theorem 3.4, i.e. we show that the bootstrap test has asymptotic level α as specified in (3.16). It follows from (3.13) that in the case $\hat{d}_2 = d_2(\hat{\beta}_1, \hat{\beta}_2) \geq \varepsilon$ the constrained estimators $\hat{\hat{\beta}}_1$ and $\hat{\hat{\beta}}_2$ coincide with the unconstrained

OLS-estimators $\hat{\beta}_1$ and $\hat{\beta}_2$, respectively. This yields in particular $\hat{\hat{d}}_2 = d_2(\hat{\hat{\beta}}_1, \hat{\hat{\beta}}_2) = \hat{d}_2$ whenever $\hat{d}_2 \geq \varepsilon$.
If $d_2 > \varepsilon$ we have

$$\begin{aligned}
\mathbb{P}(\hat{d}_2 < \hat{q}_\alpha) &= \mathbb{P}(\hat{d}_2 < \hat{q}_\alpha,\ \hat{d}_2 \geq \varepsilon) + \mathbb{P}(\hat{d}_2 < \hat{q}_\alpha,\ \hat{d}_2 < \varepsilon) \\
&\leq \mathbb{P}(\hat{d}_2 < \hat{q}_\alpha,\ \hat{\hat{d}}_2 = \hat{d}_2) + \mathbb{P}(\hat{d}_2 < \varepsilon) \\
&\leq \mathbb{P}(\hat{\hat{d}}_2 < \hat{q}_\alpha) + \mathbb{P}\Big(\frac{\sqrt{n}(\hat{d}_2 - d_2)}{\sigma_{d_2}} < \frac{\sqrt{n}(\varepsilon - d_2)}{\sigma_{d_2}}\Big).
\end{aligned}$$

Observing that $\varepsilon - d_2 < 0$, it now follows from Theorem 3.1 that the second term is of order $o(1)$. On the other hand, we have from (6.18) that the first term is of the same order, which gives $\lim_{n_1,n_2\to\infty} \mathbb{P}(\hat{d}_2 < \hat{q}_\alpha) = 0$ and proves the first part of Theorem 3.4 in the case $d_2 > \varepsilon$.
For a proof of the corresponding statement in the case $d_2 = \varepsilon$ we note that it follows again from (6.18)

$$\begin{aligned}
\mathbb{P}(\hat{d}_2 < \hat{q}_\alpha) &= \mathbb{P}(\hat{d}_2 < \hat{q}_\alpha,\ \hat{d}_2 \geq \varepsilon) + \mathbb{P}(\hat{d}_2 < \hat{q}_\alpha,\ \hat{d}_2 < \varepsilon) \\
&= \mathbb{P}(\hat{d}_2 < \hat{q}_\alpha,\ \hat{\hat{d}}_2 = \hat{d}_2) + \mathbb{P}(\hat{d}_2 < \hat{q}_\alpha,\ \hat{\hat{d}}_2 = \varepsilon) - \mathbb{P}(\hat{d}_2 < \hat{q}_\alpha, \hat{d}_2 = \varepsilon) \\
&= \mathbb{P}(\hat{d}_2 < \hat{q}_\alpha,\ \hat{\hat{d}}_2 = \hat{d}_2) + \mathbb{P}(\hat{d}_2 < \hat{q}_\alpha,\ \hat{\hat{d}}_2 = \varepsilon) + o(1) \\
&\overset{d_2=\varepsilon}{=} \mathbb{P}\Big(\frac{\sqrt{n}(\hat{d}_2 - d_2)}{\sigma_{d_2}} < \frac{\sqrt{n}(\hat{q}_\alpha - \hat{\hat{d}}_2)}{\sigma_{d_2}}, \hat{\hat{d}}_2 = \varepsilon\Big) + o(1) \\
&= \mathbb{P}\Big(\frac{\sqrt{n}(\hat{d}_2 - d_2)}{\sigma_{d_2}} < \frac{\sqrt{n}(\hat{q}_\alpha - \hat{\hat{d}}_2)}{\sigma_{d_2}}\Big) \\
&\qquad - \mathbb{P}\big(\hat{d}_2 - d_2 < \hat{q}_\alpha - \hat{\hat{d}}_2,\ \hat{\hat{d}}_2 > \varepsilon\big) + o(1), \qquad (6.19)
\end{aligned}$$

where the third equality is a consequence of the fact that $\sqrt{n}(\hat{d}_2 - d_2)$ is asymptotically normal distributed, which gives

$$\mathbb{P}(\hat{d}_2 < \hat{q}_\alpha, \hat{d}_2 = \varepsilon) \leq \mathbb{P}(\hat{d}_2 = \varepsilon) \longrightarrow 0.$$

If $\hat{\hat{d}}_2 > \varepsilon$ it follows that $\hat{d}_2 = \hat{\hat{d}}_2 > \varepsilon = d_2$ and consequently the second term in (6.19) can be bounded by (observing again (6.18))

$$\mathbb{P}(\hat{d}_2 - d_2 < \hat{q}_\alpha - \hat{\hat{d}}_2, \hat{d}_2 > \varepsilon) \leq \mathbb{P}(\hat{q}_\alpha - \hat{\hat{d}}_2 > 0) = o(1).$$

Therefore we obtain from Theorem 3.1, (6.17) and (6.19) that $\lim_{n\to\infty} \mathbb{P}(\hat{d}_2 < \hat{q}_\alpha) = \Phi(u_\alpha) = \alpha$, which completes the proof of part (1) of Theorem 3.4.

Proof of (2). Finally, we consider the case $d_2 < \varepsilon$ and show the consistency of the test (3.16). Theorem 3.1 implies that $\hat{d}_2 \overset{\mathbb{P}}{\longrightarrow} d_2$. Since $d_2 < \varepsilon$, there exists a constant $\delta > 0$ such that $\mathbb{P}(\hat{d}_2 < \varepsilon - \delta) \to 1$. Hence the assertion will follow if we establish that $\mathbb{P}(\hat{q}_\alpha > \varepsilon - \delta) \to 1$. To show this, denote by $F^{(n)}_{b_1,b_2,s_1,s_2}$ the conditional distribution function of $\hat{d}_2^*$ given $\hat{\hat{\beta}}_\ell = b_\ell, \hat{\sigma}_\ell = s_\ell, \ell = 1,2$. Since $\mathbb{P}(\max_{\ell=1,2} |\hat{\sigma}_\ell - \sigma_\ell| \le r) \to 1$ for any $r > 0$, it suffices to establish that for some $r > 0$

$$\sup\big\{F^{(n)}_{b_1,b_2,s_1,s_2}(\varepsilon - \delta) \mid b_\ell \in B_\ell, \ell = 1,2; \max_{\ell=1,2} |s_\ell - \sigma_\ell| \le r\big\} \to 0.$$

By uniform continuity of the map $(b_1, b_2) \mapsto d_2(b_1, b_2)$ it suffices to prove that for all $\eta > 0$ and $\ell = 1,2$

$$\sup\Big\{\mathbb{P}(|\hat{\beta}^*_\ell - b_\ell| \ge \eta \mid \hat{\hat{\beta}}_k = b_k, \hat{\sigma}_k = s_k, k = 1,2)\Big|\ (b_\ell, s_\ell) : |s_\ell - \sigma_\ell| \le r, b_\ell \in B_\ell, \ell = 1,2\Big\} \to 0. \tag{6.20}$$

We will only prove the above statement for $\ell = 1$ since the case $\ell = 2$ follows by exactly the same arguments. For $i = 1, ..., k_1, j = 1, ..., n_{1,i}$ let $e_{i,j}$ i.i.d. $\sim \mathcal{N}(0,1)$ and define

$$\begin{aligned}
\psi^{(n)}_{a,1}(b) &:= \sum_{i=1}^{k_1} \frac{n_{1,i}}{n_1}(m_1(x_{1,i}, a) - m_1(x_{1,i}, b))^2, \\
\hat{\gamma}_s &:= \frac{1}{n_1}\sum_{i=1}^{k_1}\sum_{j=1}^{n_{1,i}}(se_{i,j})^2 \\
\hat{\psi}_{a,s}(b) &:= \frac{1}{n_1}\sum_{i=1}^{k_1}\sum_{j=1}^{n_{1,i}}(m_1(x_{1,i}, a) + se_{i,j} - m_1(x_{1,i}, b))^2.
\end{aligned}$$

By construction, the conditional distribution of $\hat{\beta}_1^*$ given $\hat{\hat{\beta}}_1 = a, \hat{\sigma}_1 = s$ is equal to the distribution of the random variable $\hat{b}_{a,s} := \arg\min_{b\in B_1} \hat{\psi}_{a,s}(b)$. On the other hand, $\arg\min_{b\in B_1} \hat{\psi}_{a,s}(b) = \arg\min_{b\in B}(\hat{\psi}_{a,s}(b) - \hat{\gamma}_s)$, and

$$\hat{\psi}_{a,s}(b) - \hat{\gamma}_s = \psi^{(n)}_{a,1}(b) + 2s\sum_{i=1}^{k_1}(m_1(x_{1,i}, a) - m_1(x_{1,i}, b))\frac{1}{n_1}\sum_{j=1}^{n_{1,i}} e_{i,j}.$$

Observing that the terms $|m_1(x_{1,i}, a) - m_1(x_{1,i}, b)|$ are uniformly bounded (with respect to $a, b \in B_1$ and $x_{1,i} \in \mathcal{X}$) it follows that

$$R_n := \sup_{|s-\sigma_\ell|\le r}\sup_{a,b\in B_1}\Big|2s\sum_{i=1}^{k_1}(m_1(x_{1,i}, a) - m_1(x_{1,i}, b))\frac{1}{n_1}\sum_{j=1}^{n_{1,i}} e_{i,j}\Big| = o_{\mathbb{P}}(1)$$

since $\max_{i=1,...,k_1} |\frac{1}{n_1}\sum_{j=1}^{n_{1,i}} e_{i,j}| = o_{\mathbb{P}}(1)$. Now we obtain from Assumption 6.5 that, for sufficiently large n,

$$\sup_{(a,s):|s-\sigma_\ell|\le r} \mathbb{P}(|\hat{\beta}^*_\ell - a| \ge \eta | \hat{\hat{\beta}}_\ell = a, \hat{\sigma}_\ell = s) \le \sup_{(a,s):|s-\sigma_j|\le r} \mathbb{P}(|\hat{b}_{a,s} - a| \ge \eta)\ \le\ \mathbb{P}(R_n \ge v_\eta/4) = o(1).$$

Thus (6.20) follows, which completes the proof of Theorem 3.4. □

6.3 Proof of Theorem 4.1

Recall the definition of the estimate $\hat{d}_\infty$ in (4.1) and define the random variables

$$D_n = \sqrt{n}\,(\hat{d}_\infty - d_\infty) = \sqrt{n}\,\big(\max_{x\in\mathcal{X}} |m_1(x,\hat{\beta}_1) - m_2(x,\hat{\beta}_2)| - d_\infty\big), \tag{6.21}$$

$$Z_n = \sqrt{n}\,\big(\max_{x\in\mathcal{E}} |m_1(x,\hat{\beta}_1) - m_2(x,\hat{\beta}_2)| - d_\infty\big). \tag{6.22}$$

We will use similar arguments as given in Raghavachari (1973) and show in the following discussion that

$$R_n = D_n - Z_n = o_{\mathbb{P}}(1), \tag{6.23}$$

$$Z_n \xrightarrow{\mathcal{D}} \mathcal{Z}, \tag{6.24}$$

which proves the assertion of Theorem 4.1. For a proof of (6.23) we recall the definition of the "true" difference $\Delta(x,\beta_1,\beta_2)$ in (2.6) and the definition of the process $\{p_n(x)\}_{x\in\mathcal{X}}$ in (6.6). It follows from (6.7) and the continuous mapping theorem that

$$\lim_{n_1,n_2\to\infty} \mathbb{P}\big(\max_{x\in\mathcal{X}} |p_n(x)| > a_n\big) = 0 \tag{6.25}$$

as $n_1, n_2 \to \infty$, $n/n_1 \to \lambda \in (1,\infty)$, where $a_n = \log n/\sqrt{n}$. By the representation $p_n(x) = G_n(x) + o_{\mathbb{P}}(n^{-1/2})$ uniformly in $x \in \mathcal{X}$ and the definition of G_n in (6.6) we have for every $\eta > 0$

$$\lim_{\delta\downarrow 0} \lim_{n_1,n_2\to\infty} \mathbb{P}\Big(\sup_{\|x-y\|<\delta} \sqrt{n}|p_n(x) - p_n(y)| > \eta\Big) = 0, \tag{6.26}$$

where $\|\cdot\|$ denotes a norm on $\mathcal{X} \subset \mathbb{R}^d$. In the following discussion define the sets

$$\mathcal{E}_n^{\mp} = \big\{x \in \mathcal{X} \mid |\mp d_\infty - \Delta(x,\beta_1,\beta_2)| \le a_n\big\} \tag{6.27}$$

and $\mathcal{E}_n = \mathcal{E}_n^+ \cup \mathcal{E}_n^-$, then it follows from the definition of R_n and (6.25) that

$$\begin{aligned} 0 \le R_n &= \sqrt{n}\Big(\max_{x\in\mathcal{X}} |\Delta(x,\hat{\beta}_1,\hat{\beta}_2)| - \max_{x\in\mathcal{E}} |\Delta(x,\hat{\beta}_1,\hat{\beta}_2)|\Big) \\ &= \sqrt{n}\Big(\max_{x\in\mathcal{E}_n} |\Delta(x,\hat{\beta}_1,\hat{\beta}_2)| - \max_{x\in\mathcal{E}} |\Delta(x,\hat{\beta}_1,\hat{\beta}_2)|\Big) + o_{\mathbb{P}}(1) \\ &\le \max(R_n^-, R_n^+) + o_{\mathbb{P}}(1), \end{aligned}$$

where the quantities R_n^- and R_n^+ are defined by

$$R_n^{\mp} = \sqrt{n}\Big(\max_{x\in\mathcal{E}_n^{\mp}} |\Delta(x,\hat{\beta}_1,\hat{\beta}_2)| - \max_{x\in\mathcal{E}^{\mp}} |\Delta(x,\hat{\beta}_1,\hat{\beta}_2)|\Big).$$

We now prove the estimate $R_n^{\mp} = o_{\mathbb{P}}(1)$, which completes the proof of assertion (6.23). For this purpose we restrict ourselves to the random variable R_n^+ (the assertion for R_n^- is obtained by similar arguments). Note that $\mathcal{E}^+ \subset \mathcal{E}_n^+$ and therefore it follows that

$$
\begin{aligned}
0 \le R_n^+ &= \sqrt{n}\Big(\max_{x\in\mathcal{E}_n^+} \Delta(x,\hat\beta_1,\hat\beta_2) - \max_{x\in\mathcal{E}^+} \Delta(x,\hat\beta_1,\hat\beta_2)\Big) + o_{\mathbb{P}}(1) \qquad (6.28)\\
&\le \max_{x\in\mathcal{E}_n^+} \sqrt{n} p_n(x) - \max_{x\in\mathcal{E}^+} \sqrt{n} p_n(x) + \sqrt{n}\Big\{ \max_{x\in\mathcal{E}_n^+} \Delta(x,\beta_1,\beta_2) - d_\infty \Big\} + o_{\mathbb{P}}(1)\\
&= \max_{x\in\mathcal{E}_n^+} \sqrt{n} p_n(x) - \max_{x\in\mathcal{E}^+} \sqrt{n} p_n(x) + o_{\mathbb{P}}(1).
\end{aligned}
$$

Now define for $\gamma > 0$ the set

$$
\mathcal{E}^+(\gamma) = \{x \in \mathcal{X} \mid \exists\; y \in \mathcal{E}^+ \quad \text{with} \quad \|x-y\| < \gamma\}
$$

and $\delta_n = 2\inf\{\gamma > 0 \mid \mathcal{E}_n^+ \subset \mathcal{E}^+(\gamma)\}$. Obviously $\mathcal{E}_n^+ \subset \mathcal{E}^+(\delta_n)$ and the sequence $(\delta_n)_{n\in\mathbb{N}}$ is decreasing, such that $\delta := \lim_{n\to\infty} \delta_n$ exists. By the definition of δ_n we have $\mathcal{E}_n^+ \not\subset \mathcal{E}^+(\delta_n/4)$. Consequently, there exist $x_n \in \mathcal{E}_n^+ \subset \mathcal{X}$ such that $\|x_n - y\| \ge \delta_n/4$ for all $y \in \mathcal{E}^+$ and all $n \in \mathbb{N}$. The sequence (x_n) contains a convergent subsequence (because $\mathcal{X}$ is compact), say $(x_{n_k})_{k\in\mathbb{N}}$ which satisfies

$$
\lim_{k\to\infty} x_{n_k} = x \in \mathcal{X}, \quad d_\infty = \lim_{k\to\infty} \Delta(x_{n_k},\beta_1,\beta_2) = \Delta(x,\beta_1,\beta_2).
$$

Consequently, $x \in \mathcal{E}^+$, but by construction $\|x_{n_k} - x\| \ge \delta_n/4$ for all $k \in \mathbb{N}$, which is only possible if $\delta = \lim_{n\to\infty} \delta_n = 0$.
Now it follows from inequality (6.28) for the sequence $(\delta_n)_{n\in\mathbb{N}}$

$$
\begin{aligned}
o_{\mathbb{P}}(1) \le R_n^+ &\le \max_{x\in\mathcal{E}^+(\delta_n)} \sqrt{n} p_n(x) - \max_{x\in\mathcal{E}^+} \sqrt{n} p_n(x) + o_{\mathbb{P}}(1)\\
&\le \max_{\|y-x\|\le\delta_n} \sqrt{n}|p_n(x) - p_n(y)| + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1),
\end{aligned}
$$

where the last estimate is a consequence of (6.26). A similar statement for R_n^- completes the proof of (6.23).
For a proof of the second assertion (6.24) we define the random variable

$$
\tilde Z_n = \max\big\{ \max_{x\in\mathcal{E}^+} \sqrt{n} p_n(x);\; \max_{x\in\mathcal{E}^-}(-\sqrt{n} p_n(x))\big\},
$$

then it follows from (6.7) and the continuous mapping theorem that $\tilde Z_n \xrightarrow{\mathcal{D}} \mathcal{Z}$, where the random variable $\mathcal{Z}$ is defined in (4.4). Observing the uniform convergence in (6.25) we have as $n_1, n_2 \to \infty$

$$
\begin{aligned}
\mathbb{P}(Z_n \le t) &= \mathbb{P}\big(Z_n \le t,\; \max_{x\in\mathcal{X}} |p_n(x)| < \tfrac{d_\infty}{2}\big) + o(1) = \mathbb{P}\big(\tilde Z_n \le t,\; \max_{x\in\mathcal{X}} |p_n(x)| < \tfrac{d_\infty}{2}\big) + o(1)\\
&= \mathbb{P}(\tilde Z_n \le t) + o(1) = \mathbb{P}(\mathcal{Z} \le t) + o(1).
\end{aligned}
$$

This proves the remaining statement and completes the proof of Theorem 4.1. □

6.4 Proof of Theorem 4.2

Throughout this proof, let $\hat{\hat{\beta}}_1$ and $\hat{\hat{\beta}}_2$ denote the estimators defined in (3.13). Similarly to the proof of Theorem 3.4, it is possible to establish that

$$\hat{\hat{\beta}}_\ell \xrightarrow{\mathbb{P}} \beta_\ell \quad \ell = 1, 2, \quad \text{whenever} \quad d_\infty \geq \varepsilon, \tag{6.29}$$

$$\big\{\sqrt{n} p_n^*(x)\big\}_{x\in\mathcal{X}} \xrightarrow{\mathcal{D}} \{G(x)\}_{x\in\mathcal{X}} \tag{6.30}$$

$$p_n^*(x) = G_n^*(x) + o_{\mathbb{P}}(n^{-1/2})$$

uniformly with respect to $x \in \mathcal{X}$ where p_n^*, G_n^* are defined as in (6.12), (6.13), respectively and $G(x)$ denotes the Gaussian process defined in (6.5). As $\#\mathcal{E} = 1$, we can assume without loss of generality that $\mathcal{E} = \mathcal{E}^+ = \{x_0\}$. This gives

$$d_\infty = \max_{x\in\mathcal{X}} |m_1(x, \beta_1) - m_2(x, \beta_2)| = m_1(x_0, \beta_1) - m_2(x_0, \beta_2),$$

and Theorem 4.1 yields

$$\sqrt{n}\ (\hat{d}_\infty - d_\infty) \xrightarrow{D} G(x_0). \tag{6.31}$$

We begin with a proof of (4.9), i.e. the null hypothesis $d_\infty \geq \varepsilon$ is satisfied, and define $\mathcal{F}_n^* = \mathcal{F}_n^{+*} \cup \mathcal{F}_n^{-*}$, where

$$\mathcal{F}_n^{\mp *} = \big\{x \in \mathcal{X} \mid m_1(x, \hat{\hat{\beta}}_1) - m_2(x, \hat{\hat{\beta}}_2) = \mp \hat{\hat{d}}_\infty\big\}.$$

From (6.29) and the continuous mapping theorem we obtain the existence of a sequence $(\gamma_n)_{\in\mathbb{N}}$ such that $\gamma_n \to 0$ and

$$\sup_{x\in\mathcal{X}} |\Delta(x, \hat{\hat{\beta}}_1, \hat{\hat{\beta}}_2) - \Delta(x, \beta_1, \beta_2)| = o_{\mathbb{P}}(\gamma_n),\ \sup_{x\in\mathcal{X}} |\Delta(x, \hat{\beta}_1^*, \hat{\beta}_2^*) - \Delta(x, \hat{\hat{\beta}}_1, \hat{\hat{\beta}}_2)| = o_{\mathbb{P}}(a_n), \tag{6.32}$$

where $a_n = \log n/\sqrt{n}$ and the second statement follows from (6.30). Moreover, from the representation $p_n^* = G_n^* + o_{\mathbb{P}}(n^{-1/2})$ we have for every $\eta > 0$

$$\lim_{\delta\downarrow 0} \lim_{n_1,n_2\to\infty} \mathbb{P}\Big(\sqrt{n} \sup_{\|x-y\|<\delta} |p_n^*(x) - p_n^*(y)| > \eta\Big) = 0. \tag{6.33}$$

Now define $b_n = \max\{\gamma_n, a_n\}$ and consider the sets

$$\mathcal{F}_n^\pm = \{x \in \mathcal{X} \mid | \pm d_\infty - \Delta(x, \beta_1, \beta_2)| \leq b_n\}$$

and $\mathcal{F}_n = \mathcal{F}_n^+ \cup \mathcal{F}_n^-$. Defining the random variables

$$D_n^* = \sqrt{n}\ (\hat{d}_\infty^* - \hat{\hat{d}}_\infty) = \sqrt{n}\ \big(\max_{x\in\mathcal{X}} |m_1(x, \hat{\beta}_1^*) - m_2(x, \hat{\beta}_2^*)| - \hat{\hat{d}}_\infty\big),$$

$$Z_n^* = \sqrt{n}\ \big(\max_{x\in\mathcal{F}_n^*} |m_1(x, \hat{\beta}_1^*) - m_2(x, \hat{\beta}_2^*)| - \hat{\hat{d}}_\infty\big),$$

and observing (6.32) yields the estimate

$$\begin{aligned} 0 \le R_n^* &= D_n^* - Z_n^* = \sqrt{n}\Big(\max_{x\in\mathcal{X}} |\Delta(x,\hat\beta_1^*,\hat\beta_2^*)| - \max_{x\in\mathcal{F}_n^*} |\Delta(x,\hat\beta_1^*,\hat\beta_2^*)|\Big) \\ &= \sqrt{n}\Big(\max_{x\in\mathcal{F}_n} |\Delta(x,\hat\beta_1^*,\hat\beta_2^*)| - \max_{x\in\mathcal{F}_n^*} |\Delta(x,\hat\beta_1^*,\hat\beta_2^*)|\Big) + o_{\mathbb{P}}(1) \\ &\le \max(R_n^{-^*}, R_n^{+^*}) + o_{\mathbb{P}}(1), \end{aligned}$$

where the quantities $R_n^{-^*}$ and $R_n^{+^*}$ are defined by

$$R_n^{\mp^*} = \sqrt{n}\Big(\max_{x\in\mathcal{F}_n^{\mp}} |\Delta(x,\hat\beta_1^*,\hat\beta_2^*)| - \max_{x\in\mathcal{F}_n^{\mp^*}} |\Delta(x,\hat\beta_1^*,\hat\beta_2^*)|\Big),$$

and we define the maximum of the empty set as zero. We now prove the estimate $R_n^{+^*} = o_{\mathbb{P}}(1)$. By the definition of the set $\mathcal{F}_n^+$ we have

$$\sup_{x\notin\mathcal{F}_n^+} \Delta(x,\hat{\hat\beta}_1,\hat{\hat\beta}_2) - \Delta(x_0,\hat{\hat\beta}_1,\hat{\hat\beta}_2) \le \sup_{x\notin\mathcal{F}_n^+} \Delta(x,\beta_1,\beta_2) - \Delta(x_0,\beta_1,\beta_2) + o_{\mathbb{P}}(b_n) < -b_n + o_{\mathbb{P}}(b_n).$$

Consequently, $\mathbb{P}\left(\mathcal{F}_n^{+*} \subset \mathcal{F}_n^+\right) \longrightarrow 1$ and it follows

$$\mathbb{P}\Big(\max_{x\in\mathcal{F}_n^+} \Delta(x,\hat{\hat\beta}_1,\hat{\hat\beta}_2) = \hat{\hat d}_\infty\Big) \longrightarrow 1.$$

Therefore $0 \le R_n^{+^*} + o_{\mathbb{P}}(1)$ and

$$\begin{aligned} o_{\mathbb{P}}(1) \le R_n^{+^*} &= \sqrt{n}\Big(\max_{x\in\mathcal{F}_n^+} \Delta(x,\hat\beta_1^*,\hat\beta_2^*) - \max_{x\in\mathcal{F}_n^{+*}} \Delta(x,\hat\beta_1^*,\hat\beta_2^*)\Big) + o_{\mathbb{P}}(1) \\ &\le \max_{x\in\mathcal{F}_n^+} \sqrt{n}p_n^*(x) - \max_{x\in\mathcal{F}_n^{+*}} \sqrt{n}p_n^*(x) + \sqrt{n}\Big\{\max_{x\in\mathcal{F}_n^+} \Delta(x,\hat{\hat\beta}_1,\hat{\hat\beta}_2) - \hat{\hat d}_\infty\Big\} + o_{\mathbb{P}}(1) \\ &= \max_{x\in\mathcal{F}_n^+} \sqrt{n}p_n^*(x) - \max_{x\in\mathcal{F}_n^{+*}} \sqrt{n}p_n^*(x) + o_{\mathbb{P}}(1). \end{aligned}$$

Now we construct a set $\mathcal{F}^*(\delta_n)$ containing the point x_0 and the set $\mathcal{F}_n^+$. For this purpose consider the ball with center x_0 and radius γ

$$\mathcal{F}^*(\gamma) := \{x \in \mathcal{X} \mid \|x - x_0\| < \gamma\}$$

and define $\delta_n = 2 \cdot \inf\{\gamma > 0 \mid \mathcal{F}_n^+ \subset \mathcal{F}^*(\gamma)\}$. Obviously $x_0 \in \mathcal{F}^*(\delta_n)$. Moreover, without loss of generality we assume that the sequence b_n is decreasing. As a consequence $(\delta_n)_{n\in\mathbb{N}}$ is also decreasing, such that $\delta := \lim_{n\to\infty}\delta_n$ exists. By the definition of δ_n we have $\mathcal{F}_n^+ \not\subset \mathcal{F}^*(\delta_n/4)$. Consequently, there exists an $x_n \in \mathcal{F}_n^+$ such that $\|x_n - x_0\| \ge \delta_n/4$. As $\mathcal{X}$ is compact, there exists a convergent subsequence, say $(x_{n_k})_{k\in\mathbb{N}}$ with limit $\lim_{k\to\infty} x_{n_k} = x \in \mathcal{X}$ and

$$d_\infty = \lim_{n\to\infty} \Delta(x_{n_k},\beta_1,\beta_2) = \Delta(x,\beta_1,\beta_2).$$

Consequently $x = x_0$, and from $\|x_{n_k} - x_0\| \geq \delta/4$ for all $k \in \mathbb{N}$ we obtain $\delta = \lim_{n\to\infty} \delta_n = 0$. Note that $\mathcal{F}_n^{+*} \subset \mathcal{F}_n^{+} \subset \mathcal{F}^*(\delta_n)$ with probability tending to one, which yields

$$\begin{aligned} o_{\mathbb{P}}(1) \leq R_n^{+^*} &\leq \max_{x\in\mathcal{F}_n^+} \sqrt{n} p_n^*(x) - \max_{x\in\mathcal{F}_n^{+*}} \sqrt{n} p_n^*(x) + o_{\mathbb{P}}(1) \\ &\leq \sqrt{n} \max_{x,y\in\mathcal{F}^*(\delta_n)} |p_n^*(x) - p_n^*(y)| + o_{\mathbb{P}}(1) \\ &\leq \max_{\|x-y\|<2\delta_n} \sqrt{n}|p_n^*(x) - p_n^*(y)| + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1) \end{aligned}$$

where the last equality follows from (6.33). Moreover, since we assumed that $\mathcal{E} = \mathcal{E}^+$ it can be shown that $\mathcal{F}_n^-, \mathcal{F}_n^{-*}$ will be empty with probability tending to one and thus $R_n^* = o_{\mathbb{P}}(1)$. Now, by (6.32), (6.33) and the estimate $\sup_{x\in\mathcal{F}_n^{+*}} \|x - x_0\| = o_{\mathbb{P}}(1)$ we have

$$\tilde{Z}_n^* := \sqrt{n} \max\Big\{ \max_{x\in\mathcal{F}_n^{+*}} p_n^*(x), \max_{x\in\mathcal{F}_n^{-*}} (-p_n^*(x))\Big\} = \sqrt{n} p_n^*(x_0) + o_{\mathbb{P}}(1) \xrightarrow{\mathcal{D}} G(x_0),$$

and it follows as in the previous proof that $Z_n^* = \tilde{Z}_n^* + o_{\mathbb{P}}(1)$, which gives $Z_n^* \xrightarrow{\mathcal{D}} G(x_0)$ conditionally on χ in probability. This finally yields

$$\sqrt{n}(\hat{d}_\infty^* - \hat{\hat{d}}_\infty) \xrightarrow{\mathcal{D}} G(x_0)$$

conditionally on χ in probability.

As $\hat{q}_{\alpha,\infty}$ is the α-quantile of the bootstrap test statistics $\hat{d}_\infty^*$ conditionally on χ it holds that

$$\alpha = \mathbb{P}\big(\hat{d}_\infty^* < \hat{q}_{\alpha,\infty}\big|\chi\big) = \mathbb{P}\big(\sqrt{n}(\hat{d}_\infty^* - \hat{\hat{d}}_\infty) < \sqrt{n}(\hat{q}_{\alpha,\infty} - \hat{\hat{d}}_\infty)\big|\chi\big),$$

almost surely. Thus the α-quantile of the distribution of $\sqrt{n}(\hat{d}_\infty^* - \hat{\hat{d}}_\infty)$ conditionally on χ is of the form $\hat{p}_{\alpha,\infty} := \sqrt{n}(\hat{q}_{\alpha,\infty} - \hat{\hat{d}}_\infty)$ and satisfies $\hat{p}_{\alpha,\infty} \xrightarrow{\mathbb{P}} z_\alpha$, where z_α denotes the α-quantile of the distribution of $G(x_0)$. With $\sigma_{d_\infty}^2 := \mathrm{Var}(G(x_0))$, it now follows from Lemma 21.2 in Van der Vaart (1998)

$$\frac{\hat{p}_\alpha}{\sigma_{d_\infty}} \xrightarrow{\mathbb{P}} u_\alpha,$$

where u_α denotes the α-quantile of the standard normal distribution. This result is the analogue of (6.17) in the proof of Theorem 3.4, and the assertion now follows by exactly the same arguments as given in the proof of this result.

Finally, for a proof of the remaining statement (3.18) we note that the map $(b_1, b_2) \mapsto d_\infty(b_1, b_2)$ is uniformly continuous. Therefore, the result follows by exactly the same arguments as given in the proof of Theorem 3.4. The details are omitted for the sake of brevity. □