

BOOTSTRAPPING MANSKI'S MAXIMUM SCORE ESTIMATOR

Emilio Seijo and Bodhisattva Sen
Columbia University

Abstract

In this paper we study the consistency of different bootstrap methods for making inference on Manski's maximum score estimator. We propose three new, model-based bootstrap procedures for this problem and prove their consistency. Simulation experiments are carried out to evaluate their performance and to compare them with the m -out-of- n bootstrap method. Our results indicate that one of our proposed methods, the smooth bootstrap (see Section 2.2.4), outperforms all the others. Moreover, we also conclude that the other two proposed procedures, fixed and random design residual bootstraps (see Sections 2.2.2 and 2.2.3, respectively), perform at least as well as the m -out-of- n bootstrap procedure while being less dependent on the choice of tuning parameter(s). We provide a set of sufficient conditions for the consistency of any bootstrap method in this problem. Additionally, we prove a convergence theorem for triangular arrays of random variables arising from binary choice models, which may be of independent interest.

1 Introduction

Consider a (latent-variable) binary response model of the form

$$Y = \mathbf{1}_{\beta_0^T X + U \geq 0},$$

where $\mathbf{1}$ is the indicator function, X is an \mathbb{R}^d -valued, continuous random vector of explanatory variables, U is an unobserved random variable and $\beta_0 \in \mathbb{R}^d$ is an unknown vector with $|\beta_0| = 1$. The parameter of interest is β_0 . If the conditional distribution of U given X is known up to a finite set of parameters, maximum likelihood techniques can be used for estimation, among other methods; see, for example, [McFadden](#)

(1974). The parametric assumption on U may be relaxed in several ways. For instance, if U and X are independent or if the distribution of U depends on X only through the index $\beta_0^T X$, the semiparametric estimators of Han (1987), Horowitz and Härdle (1996), Powell and Stoker (1989), and Sherman (1993) can be used; also see Cosslett (1983). The maximum score estimator considered by Manski (1975) permits the distribution of U to depend on X in an unknown and very general way (heteroscedasticity of unknown form). The model replaced parametric assumptions on the error disturbance U with a conditional median restriction, i.e., $\mathbf{med}(U|X) = 0$, where $\mathbf{med}(U|X)$ represents the conditional median of U given X . Given n observations $(X_1, Y_1), \dots, (X_n, Y_n)$ from such a model, Manski (1975) defined a *maximum score estimator* as any maximizer of the objective function $\sum_{j=1}^n (Y_j - \frac{1}{2}) \mathbf{1}_{\beta^T X_j \geq 0}$ over the unit sphere in \mathbb{R}^d .

The asymptotics of the maximum score estimator are well-known. Under some regularity conditions, the estimator was shown to be strongly consistent in Manski (1985) and its asymptotic distribution was derived in Kim and Pollard (1990) (also see Cavanagh (1987)). However, the complicated nature of its limit law (which depends, among other parameters, on the conditional distribution of U given X for values of X on the hyperplane $\{x \in \mathbb{R}^d : \beta_0^T x = 0\}$) and the fact that it exhibits nonstandard asymptotics (cube-root rate of convergence) have made it difficult to do inference for the maximum score estimator under the complete generality of the model. Hence, many authors have proposed different estimation and inference procedures that work under stronger assumptions on the conditional distribution of U given X ; see e.g., Horowitz (1992) and Horowitz (2002).

In this context, the bootstrap stands out as a natural method for inference. Bootstrap methods avoid the problem of estimating nuisance parameters and are generally reliable in problems with \sqrt{n} convergence rates; see Bickel and Freedman (1981), Singh (1981), Shao and Tu (1995) and its references. Unfortunately, the classical bootstrap (drawing n observations with replacement from the original data) is *inconsistent* for the maximum score estimator as shown in Abrevaya and Huang (2005). In

fact, the bootstrap estimator can behave quite erratically in cube-root convergence problems. For instance, it was shown in [Sen et al. \(2010\)](#) that for the Grenander estimator (the nonparametric maximum likelihood estimator of a non-increasing density on $[0, \infty)$), a prototypical example of cube-root asymptotics, the bootstrap estimator is not only inconsistent but has no weak limit in probability. This stronger result should also hold for the maximum score estimator (for numerical and graphical evidence see [Section 5](#) and, in particular, [Figure 1](#)). These findings contradict some of the results of [Abrevaya and Huang \(2005\)](#) (especially [Theorem 4](#) and the conclusions of [Section 4](#) of that paper) where it is claimed that for some single-parameter estimators a simple method for inference based upon the naive bootstrap can be developed in spite of its inconsistency.

Thus, in order to apply the bootstrap to this problem some modifications of the classical approach are required. Two variants of the classical bootstrap that can be applied in this situation are the so-called m -out-of- n bootstrap and subsampling. The performance of subsampling for inference on the maximum score estimator has been studied in [Delgado et al. \(2001\)](#). The consistency of the m -out-of- n bootstrap can be deduced from the results in [Lee and Pun \(2006\)](#). Despite their simplicity, the reliability of both methods depends crucially on the size of the subsample (the m in the m -out-of- n bootstrap and the block size in subsampling) and a proper choice of this tuning parameter is difficult. Thus, it would be desirable to have alternative, more automated and consistent bootstrap procedures for inference in the general setting of the binary choice model of Manski.

In this paper we propose three model-based bootstrap procedures (i.e., procedures that use the model setup and assumptions explicitly to construct bootstrap schemes; see [Section 2](#) for the details) that provide an alternative to subsampling and the m -out-of- n bootstrap methods. We prove that these procedures are consistent for the maximum score estimator under quite general assumptions. In doing so, we also prove a general convergence theorem for triangular arrays of random variables coming from binary choice models that can be used to verify the consistency of any

bootstrap scheme in this setup. We derive our results in greater generality than most authors by assuming that β_0 belongs to the unit sphere in \mathbb{R}^d as opposed to fixing its first co-ordinate to be 1 (as in [Abrevaya and Huang \(2005\)](#)). To make the final results more accessible we express them in terms of integrals with respect to the Lebesgue measure as opposed to surface measures, as in [Kim and Pollard \(1990\)](#). In addition, we run simulation experiments to compare the finite sample performances of the different bootstrap procedures. Our results indicate that one of our proposed methods, the smooth bootstrap (see [Section 2.2.4](#)), outperforms all the others. Moreover, we also conclude that the other two proposed procedures, fixed and random design residual bootstraps (see [Sections 2.2.2](#) and [2.2.3](#), respectively), perform at least as well as the best m -out-of- n bootstrap procedures while being less dependent on the choice of tuning parameters. Our analysis also illustrates the inconsistency of the classical bootstrap. To the best of our knowledge, this paper is the first attempt to understand the behavior of model-based bootstrap procedures under very general heteroscedasticity assumptions for the maximum score estimator.

Our exposition is organized as follows: in [Section 2](#) we introduce the model and our assumptions ([Section 2.1](#)) and give a brief description of bootstrap and the bootstrap procedures to be considered in this paper ([Section 2.2](#)). In [Section 3](#) we prove a general convergence theorem for triangular arrays of random variables coming from binary choice models (see [Theorem 3.1](#)) that will be useful in verifying the consistency of the proposed bootstrap schemes, investigated in [Section 4](#). We study and compare the finite sample performance of the different bootstrap schemes in [Section 5](#) through simulation experiments. Additionally, we include an Appendix ([Sections A](#) and [B](#)) with some auxiliary results and some technical details omitted from the main text.

2 The model and the bootstrap procedures

2.1 The maximum score estimator

Consider a Borel probability measure \mathbb{P} on \mathbb{R}^{d+1} , $d \geq 2$, such that if $(X, U) \sim \mathbb{P}$ then X takes values in a closed, convex region $\mathfrak{X} \subset \mathbb{R}^d$ with $\mathfrak{X}^\circ \neq \emptyset$ and U is a real-valued random variable that satisfies $\mathbf{med}(U|X) = 0$, where $\mathbf{med}(\cdot)$ represents the median. We only observe (X, Y) where $Y := \mathbf{1}_{\beta_0^T X + U \geq 0}$ for some $\beta_0 \in \mathcal{S}^{d-1}$, where \mathcal{S}^{d-1} is the unit sphere in \mathbb{R}^d (with respect to the Euclidian norm). We assume that, under \mathbb{P} , X has a continuous distribution with a strictly positive density p on \mathfrak{X}° which is continuously differentiable on \mathfrak{X}° and such that ∇p is integrable (with respect to Lebesgue measure) over \mathfrak{X} . We take the function

$$\kappa(x) := \mathbb{P}(Y = 1|X = x) = \mathbb{P}(\beta_0^T X + U \geq 0|X = x) \quad (1)$$

to be continuously differentiable on \mathfrak{X}° . Finally, we suppose that the set $\{x \in \mathfrak{X}^\circ : \nabla \kappa(x)^T \beta_0 > 0\}$ intersects the hyperplane $\{x \in \mathbb{R}^d : \beta_0^T x = 0\}$ and that $\int |\nabla \kappa(x)| x x^T p(x) dx$ is well-defined.

Observe that $Y|X = x \sim \text{Bernoulli}(\kappa(x))$ and this plays a crucial role in the sequel. Our approach is model-based and it exploits the above relationship between Y and X using a non-parametric estimator of $\kappa(x)$. The median restriction $\mathbf{med}(U|X) = 0$ on the *unobserved* variable U implies that $\beta_0^T x \geq 0$ if and only if $\kappa(x) \geq 1/2$ for all $x \in \mathfrak{X}$.

Given observations $(X_1, Y_1), \dots, (X_n, Y_n)$ from such a model, we wish to estimate $\beta_0 \in \mathcal{S}^{d-1}$. A maximum score estimator of β_0 is any element $\hat{\beta}_n \in \mathcal{S}^{d-1}$ that satisfies:

$$\hat{\beta}_n := \operatorname{argmax}_{\beta \in \mathcal{S}^{d-1}} \left\{ \frac{1}{n} \sum_{j=1}^n \left(Y_j - \frac{1}{2} \right) \mathbf{1}_{\beta^T X_j \geq 0} \right\}. \quad (2)$$

Note that there may be many elements of \mathcal{S}^{d-1} that satisfy (2). We will focus on *measurable selections* of maximum score estimators, that is, we will assume that we can compute the estimator in such a way that $\hat{\beta}_n$ is measurable (this is justified in view of the *measurable selection theorem*, see Chapter 8 of [Aubin and Frankowska \(2009\)](#)). We make this assumption to avoid the use of outer probabilities.

Our regularity conditions are similar to those in Example 6.4 of [Kim and Pollard \(1990\)](#) and imply those in [Manski \(1985\)](#). Hence, a consequence of Lemmas 2 and 3 in [Manski \(1985\)](#) is that the parameter β_0 is identifiable and the unique maximizer of the process $\Gamma(\beta) := \mathbb{P} \left[\left(Y - \frac{1}{2} \right) \mathbf{1}_{\beta^T X \geq 0} \right]$. Similarly, Theorem 1 in the same paper implies that if $(\hat{\beta}_n)_{n=1}^\infty$ is any sequence of maximum score estimators, we have $\hat{\beta}_n \xrightarrow{a.s.} \beta_0$.

2.2 The bootstrap schemes

We want to investigate the consistency of different bootstrap procedures for constructing confidence regions for β_0 . We start by briefly reviewing the bootstrap. Given a sample $\mathbf{W}_n = \{W_1, W_2, \dots, W_n\} \stackrel{\text{iid}}{\sim} L$ from an unknown distribution L on some Euclidean space (or, more generally, on some *Polish space*), suppose that the distribution function \mathbb{G}_n of some random variable $V_n \equiv V_n(\mathbf{W}_n, L)$ is of interest; V_n is usually called a *root* and it can in general be any measurable function of the data and the distribution L . The bootstrap method can be broken into three simple steps:

- (i) Construct an estimator \hat{L}_n of L from \mathbf{W}_n .
- (ii) Generate $\mathbf{W}_n^* = \{W_1^*, \dots, W_n^*\} \stackrel{\text{iid}}{\sim} \hat{L}_n$ given \mathbf{W}_n .
- (iii) Estimate \mathbb{G}_n by $\hat{\mathbb{G}}_n$, the conditional distribution of $V_n(\mathbf{W}_n^*, \hat{L}_n)$ given \mathbf{W}_n .

Let ρ denote the Prokhorov metric or any other metric metrizing weak convergence of probability measures. We say that $\hat{\mathbb{G}}_n$ is *consistent* if $\rho(\mathbb{G}_n, \hat{\mathbb{G}}_n) \xrightarrow{\mathbf{P}} 0$; if \mathbb{G}_n has a weak limit \mathbb{G} , this is equivalent to $\hat{\mathbb{G}}_n$ converging weakly to \mathbb{G} in probability. Similarly, $\hat{\mathbb{G}}_n$ is *strongly consistent* if $\rho(\mathbb{G}_n, \hat{\mathbb{G}}_n) \xrightarrow{a.s.} 0$. Intuitively, an \hat{L}_n that mimics the essential properties of the underlying distribution L can be expected to perform well. The choice of \hat{L}_n mostly considered in the literature is the empirical distribution function (EDF). Despite being a good estimator in many situations, the EDF can fail to capture some properties of L that may be crucial for the problem under consideration. This is especially true in nonstandard problems and, in particular, in the case of the maximum score estimator: it was shown in [Abrevaya and Huang \(2005\)](#) that the EDF bootstrap is inconsistent.

We will denote by $\mathcal{Z} = \sigma((X_n, Y_n)_{n=1}^\infty)$ the σ -algebra generated by the sequence $(X_n, Y_n)_{n=1}^\infty = \left(X_n, \mathbf{1}_{\beta_0^T X_n + U_n \geq 0}\right)_{n=1}^\infty$ with $(X_n, U_n)_{n=1}^\infty \stackrel{i.i.d.}{\sim} \mathbb{P}$ and write $\mathbf{P}_{\mathcal{Z}}(\cdot) = \mathbf{P}(\cdot | \mathcal{Z})$ and $\mathbf{E}_{\mathcal{Z}}(\cdot) = \mathbf{E}(\cdot | \mathcal{Z})$. We will approximate the distribution of $\Delta_n = n^{1/3}(\hat{\beta}_n - \beta_0)$ by $\mathbf{P}_{\mathcal{Z}}(\Delta_n^* \leq x)$, the conditional distribution of $\Delta_n^* = m_n^{1/3}(\beta_n^* - \hat{\beta}_n)$, and use this to build a CI for β_0 , where β_n^* is a maximum score estimator of β_0 obtained from the bootstrap sample and $\hat{\beta}_n$ is an estimator satisfying (2). In what follows we will introduce 5 bootstrap procedures to do inference on the maximum score estimator. We will prove the consistency of three of them using the results of Section 3.

To guarantee the consistency of the bootstrap procedures we need to make the following assumption on \mathbb{P} .

(A0) There is $r > 2$ and a continuous, nonnegative, integrable function $D : \mathfrak{X} \rightarrow \mathbb{R}$ such that $|\kappa(x) - 1/2| \leq D(x)|\beta_0^T x|$ for all $x \in \mathfrak{X}$ and the function $t \mapsto \mathbb{P}\left(|X|^r D(X)^r \mathbf{1}_{|\beta_0^T X| \leq t|X|}\right)$ is Lipschitz around 0.

Remark: Although condition (A0) is not common in the literature, it is indeed satisfied in many frequently encountered situations. For instance, if $\mathbb{P}(|X|^r) < \infty$ and $\nabla \kappa(X)$ is bounded on \mathfrak{X} , an application of the *mean value theorem* shows that for any $x \in \mathfrak{X}^\circ$ there is $\theta_x \in [0, 1]$ such that $|\kappa(x) - 1/2| = |\kappa(x) - \kappa(x - (\beta_0^T x)\beta_0)| = |(\beta_0^T x)\nabla \kappa(x + \theta_x(\beta_0^T x)\beta_0)^T \beta_0|$ and hence (A0) holds (with $D(x) = |\nabla \kappa(x + \theta_x(\beta_0^T x)\beta_0)|$). Thus (A0) will be true, in particular, if \mathfrak{X} is compact.

2.2.1 Scheme 1 (classical bootstrap)

Compute β_n^* from a random sample $(X_{n,1}^*, Y_{n,1}^*), \dots, (X_{n,n}^*, Y_{n,n}^*)$ from the EDF of the data $(X_1, Y_1), \dots, (X_n, Y_n)$.

2.2.2 Scheme 2 (fixed design residual bootstrap)

A natural alternative to the classical with replacement bootstrap in a regression context consists of sampling the “residuals” fixing the covariates. We will now present

a bootstrap procedure constructed in this manner. To construct the “residuals” we explicitly model $\kappa(\cdot)$, defined in (1).

- (i) Use the data $(X_1, Y_1), \dots, (X_n, Y_n)$ to find a smooth approximation $\tilde{\kappa}_n$ of κ (e.g., using kernel regression procedures) satisfying the following properties: $\tilde{\kappa}_n$ converges to κ uniformly on compact subsets of \mathfrak{X}° with probability (w.p.) 1 and there is $q > 1$ such that $\mu_n(|\tilde{\kappa}_n - \kappa|^{rq})^{1/rq} = o_{\mathbf{P}}(n^{-(r+2)/3r})$, $\mathbb{P}(|\tilde{\kappa}_n(X) - \kappa(X)|^{rq})^{1/rq} = o_{\mathbf{P}}(n^{-(r+2)/3r})$ and $\mathbb{P}\left(|X|^{\frac{rq}{q-1}}\right) < \infty$, where $r > 2$ is given by (A0) and μ_n is the empirical measure of X_1, \dots, X_n .
- (ii) Let $\hat{\kappa}_n(x) := \tilde{\kappa}_n(x)\mathbf{1}_{\hat{\beta}_n^T x(\tilde{\kappa}_n(x)-1/2) \geq 0} + \frac{1}{2}\mathbf{1}_{\hat{\beta}_n^T x(\tilde{\kappa}_n(x)-1/2) < 0}$, where $\hat{\beta}_n$ is a maximum score estimator computed from the data.
- (iii) Obtain independent random variables $Y_{n,1}^*, \dots, Y_{n,n}^*$ such that $Y_{n,j}^* \sim \text{Bernoulli}(\hat{\kappa}_n(X_j))$.
- (iv) Let β_n^* be any minimizer of

$$S_n^*(\beta) := \frac{1}{n} \sum_{k=1}^n \left(Y_{n,k}^* - \frac{1}{2} \right) \mathbf{1}_{\beta^T X_k \geq 0}$$

Remark: Steps (i) and (ii) deserve comments. We start with (ii). Note that, as opposed to $\tilde{\kappa}_n$, the modified estimator $\hat{\kappa}_n$ satisfies the inequality $(\hat{\beta}_n^T x)(\hat{\kappa}_n(x) - 1/2) \geq 0$ for all $x \in \mathfrak{X}$. This inequality guarantees that the *latent variable* structure of \mathbb{P} is reflected by the bootstrap sample. It implies the existence of independent variables $U_{n,1}^*, \dots, U_{n,n}^*$ such that $\mathbf{med}(U_{n,j}^* | X_{n,j}^*) = 0$ and $\mathbf{E}(Y_{n,j}^* | X_{n,j}^*) = \mathbf{E}\left(\mathbf{1}_{\hat{\beta}_n^T X_{n,j}^* + U_{n,j}^* \geq 0} \mid X_{n,j}^*\right)$.

With respect to (i), a reasonable question is that of whether or not the specified rates are actually achievable. According to Stone (1982) the optimal and achievable rate of convergence for estimating κ nonparametrically is $n^{-\frac{p}{2p+d}}$ if k is p times continuously differentiable over \mathfrak{X} .

2.2.3 Scheme 3 (random design residual bootstrap)

As an alternative to fixing the predictor values, we can also generate them randomly from the empirical distribution of the X 's. We describe the scheme below.

- (i) Sample $X_{n,1}^*, \dots, X_{n,n}^* \stackrel{i.i.d.}{\sim} \mu_n$, where μ_n is the empirical measure of X_1, \dots, X_n .
- (ii) Do (i)–(ii) of Scheme 2.
- (iii) Obtain independent random variables $Y_{n,1}^*, \dots, Y_{n,n}^*$ such that $Y_{n,j}^* \sim \text{Bernoulli}(\hat{\kappa}_n(X_{n,j}^*))$.
- (iv) Compute β_n^* as a maximum score estimator from the bootstrap sample $(X_{n,1}^*, Y_{n,1}^*), \dots, (X_{n,n}^*, Y_{n,n}^*)$.

2.2.4 Scheme 4 (smooth bootstrap)

It will be seen in Theorem 3.1 that the asymptotic distribution of Δ_n depends on the behavior of the distribution of X under \mathbb{P} around the hyperplane $\mathcal{H} := \{x \in \mathbb{R}^d : \beta_0^T x = 0\}$. As the EDF is a discrete distribution, a smooth approximation to $F(\cdot) := \mathbb{P}(X \in (\cdot))$, the law of X under \mathbb{P} , might yield a better finite sample approximation to the local behavior around \mathcal{H} . Indeed our simulation studies clearly illustrate this point (see Section 5). This scheme can be described as follows:

- (i) Choose an appropriate nonparametric smoothing procedure (e.g., kernel density estimation) to build a distribution \hat{F}_n with a density \hat{p}_n such that $\|\hat{F}_n - F\|_\infty \xrightarrow{a.s.} 0$ and $\hat{p}_n \rightarrow p$ pointwise on \mathfrak{X}° w.p. 1. Assume that, in addition, we have that $\|\hat{p}_n - p\|_1 = O_{\mathbf{P}}(\varepsilon_n)$ for some sequence $(\varepsilon_n)_{n=1}^\infty$ with $\varepsilon_n = o(n^{-1/3})$, where $\|\cdot\|_1$ stands for the \mathbb{L}_1 norm in \mathbb{R}^d with respect to Lebesgue measure.
- (ii) Get i.i.d. replicates $X_{n,1}^*, \dots, X_{n,n}^*$ from \hat{F}_n .
- (iii) Do steps (ii)–(iv) of Scheme 3.

2.2.5 Scheme 5 (m -out-of- n bootstrap)

Consider an increasing sequence $(m_n)_{n=1}^\infty$ with $m_n = o(n)$. Obtain a random sample of size m_n from the EDF of the data and compute the maximum score estimator β_n^* from it.

3 A convergence theorem

We will now present a convergence theorem for triangular arrays of random variables. This theorem will be applied in Section 4 to prove the consistency of the bootstrap procedures described in Sections 2.2.2, 2.2.3 and 2.2.4. We would like to point out that the results of this section hold even if assumption (A0) of Section 2.2 is not satisfied.

We start by introducing some notation. For a signed Borel measure μ on some metric space (X, ρ) and a Borel measurable function $f : X \rightarrow \mathbb{C}$ which is either integrable or nonnegative we will use the notation $\mu(f) := \int f d\mu$; if \mathcal{G} is a class of such functions on X we write $\|\mu\|_{\mathcal{G}} := \sup\{|\mu(f)| : f \in \mathcal{G}\}$. We will also make use of the sup-norm notation, that is, for functions $g : X \rightarrow \mathbb{R}^d$, $G : X \rightarrow \mathbb{R}^{d \times d}$ we write $\|g\|_X := \sup\{|g(x)| : x \in X\}$ and $\|G\|_X := \sup\{\|G(x)\|_2 : x \in X\}$, where $|\cdot|$ stands for the usual Euclidian norm and $\|\cdot\|_2$ denotes the matrix \mathbb{L}_2 -norm on the space $\mathbb{R}^{d \times d}$ of all $d \times d$ real matrices (see Meyer (2001), page 281). We will regard the elements of Euclidian spaces as column vectors.

Suppose that we are given a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ and a triangular array of random variables $\{(X_{n,j}, U_{n,j})\}_{1 \leq j \leq m_n}^{n \in \mathbb{N}}$ where $(m_n)_{n=1}^\infty$ is a sequence of natural numbers satisfying $m_n \uparrow \infty$, and $X_{n,j}$ and $U_{n,j}$ are \mathbb{R}^d and \mathbb{R} valued random variables, respectively. Furthermore, assume that the rows $\{(X_{n,1}, U_{n,1}), \dots, (X_{n,m_n}, U_{n,m_n})\}$ are formed by independent random variables. Denote by $\mathbb{Q}_{n,j}$ the distribution of $(X_{n,j}, U_{n,j})$, $1 \leq j \leq m_n$, $n \in \mathbb{N}$ and define the probability measure $\mathbb{Q}_n := \frac{1}{m_n}(\mathbb{Q}_{n,1} + \dots + \mathbb{Q}_{n,m_n})$. Consider the class of functions

$$\left\{ f_{\alpha, \beta} := \left(\mathbf{1}_{\beta^T x + u \geq 0} - \frac{1}{2} \right) \mathbf{1}_{\alpha^T x \geq 0} : \alpha, \beta \in \mathbb{R}^d \right\}$$

and the class

$$\mathcal{F} := \left\{ \left(\kappa(x) - \frac{1}{2} \right) \mathbf{1}_{\beta^T x \geq 0} : \beta \in \mathbb{R}^d \right\}.$$

Our assumptions on \mathbb{P} imply that $\Gamma(\beta) = \mathbb{P}(f_{\beta, \beta_0})$ is twice continuously differentiable on a neighborhood of β_0 (see Lemma A.1) and that the Hessian matrix $\nabla^2 \Gamma(\beta_0)$ is nonpositive definite on an open neighborhood $U \subset \mathbb{R}^d$ of β_0 . The main properties of Γ are established in Lemma A.1 of the Appendix.

We take the measures $\{\mathbb{Q}_{n,j}\}_{1 \leq j \leq m_n}^{n \in \mathbb{N}}$ to satisfy the following conditions:

- (A1) $\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{F}} \rightarrow 0$ and the sequence of x -marginals $\{\mathbb{Q}_n((\cdot) \times \mathbb{R})\}_{n=1}^{\infty}$ is uniformly tight.
- (A2) If $(X, U) \sim \mathbb{Q}_{n,j}$ with X and U taking values in \mathbb{R}^d and \mathbb{R} , respectively, then $\mathbf{med}(U|X) = 0$.
- (A3) There are $\beta_n \in \mathcal{S}^{d-1}$ and a Borel measurable function $\kappa_n : \mathfrak{X} \rightarrow [0, 1]$ such that $\kappa_n(x) = \mathbb{Q}_{n,j}(\beta_n^T X + U \geq 0 | X = x)$ for all $1 \leq j \leq m_n$ and κ_n converges to κ uniformly on compact subsets of \mathfrak{X}° . Moreover, there is $\nu > 2$ such that $\mathbb{P}(|\kappa_n(X) - \kappa(X)|^\nu)^{1/\nu} = o(m_n^{-(1+\nu)/3\nu})$.
- (A4) Let $F_{n,K}$ be a measurable envelope of the class of functions $\mathcal{F}_{n,K} := \{\mathbf{1}_{\beta^T x \geq 0} - \mathbf{1}_{\beta_n^T x \geq 0} : |\beta - \beta_n| \leq K\}$. Note that there are two small enough constants $C, K_* > 0$ such that for any $0 < K \leq K_*$ and $n \in \mathbb{N}$, $F_{n,K}$ can be taken to be of the form $\mathbf{1}_{\beta_K^T x \geq 0 > \alpha_K^T x} + \mathbf{1}_{\alpha_K^T x \geq 0 > \beta_K^T x}$ for $\alpha_K, \beta_K \in \mathbb{R}^d$ satisfying $|\alpha_K - \beta_K| \leq CK$. We assume that there exist $R_0, \Delta_0 \in (0, K_* \wedge 1]$ and a decreasing sequence $(\epsilon_n)_{n=1}^{\infty}$ of positive numbers with $\epsilon_n \downarrow 0$ such that for any $n \in \mathbb{N}$ and for any $\Delta_0 m_n^{-1/3} < R \leq R_0$ we have

- (i) $|(\mathbb{Q}_n - \mathbb{P})(F_{n,R}^2)| \leq \epsilon_1 R$;
- (ii) $\sup_{\substack{|\alpha - \beta_n| \vee |\beta - \beta_n| \leq R \\ |\alpha - \beta| \leq R}} |(\mathbb{Q}_n - \mathbb{P})(\kappa_n(X)(\mathbf{1}_{\alpha^T X \geq 0} - \mathbf{1}_{\beta^T X \geq 0}))| \leq \epsilon_n R m_n^{-1/3}$;
- (iii) $\sup_{\substack{|\alpha - \beta_n| \vee |\beta - \beta_n| \leq R \\ |\alpha - \beta| \leq R}} |(\mathbb{Q}_n - \mathbb{P})(\mathbf{1}_{\alpha^T X \geq 0} - \mathbf{1}_{\beta^T X \geq 0})| \leq \epsilon_n R m_n^{-1/3}$.

(A5) $|\nabla\Gamma(\beta_n)| = O(m_n^{-1/3})$.

In this context we write $Y_{n,j} := \mathbf{1}_{\beta_n^T X_{n,j} + U_{n,j} \geq 0}$ for $1 \leq j \leq m_n$, $M_n(\beta) := \mathbb{Q}_n(f_{\beta, \beta_n})$ and recall $\Gamma(\beta) = \mathbb{P}(f_{\beta, \beta_0})$ for $n \in \mathbb{N}$ and $\beta \in \mathcal{S}^{d-1}$. Note that the *observable* data consists of the pairs $\{(X_{n,j}, Y_{n,j})\}_{j=1}^{m_n}$. We will say that $\beta_n^* \in \mathcal{S}^{d-1}$ is a *maximum score estimator* based on $(X_{n,j}, Y_{n,j})$, $1 \leq j \leq m_n$, if it is a maximizer of $1/m_n \sum_{j=1}^{m_n} (Y_{n,j} - 1/2)\mathbf{1}_{\beta^T X_{n,j} \geq 0}$.

Before attempting to prove any asymptotic results, we will state the following lemma which establishes an important relationship between the β_n 's above and β_0 .

Lemma 3.1 *Under $\{A1-A3\}$, we have $\beta_n \rightarrow \beta_0$.*

Proof: Note that because of (A2) we can write

$$\begin{aligned} M_n(\beta) &= \frac{1}{m_n} \sum_{j=1}^{m_n} \mathbb{Q}_{n,j} \left((\mathbb{Q}_{n,j}(U \geq -\beta_n^T X | X) - \mathbb{Q}_{n,j}(U \geq 0 | X)) \mathbf{1}_{\beta^T X \geq 0, \beta_n^T X \geq 0} \right) \\ &\quad + \frac{1}{m_n} \sum_{j=1}^{m_n} \mathbb{Q}_{n,j} \left((\mathbb{Q}_{n,j}(U \geq -\beta_n^T X | X) - \mathbb{Q}_{n,j}(U \geq 0 | X)) \mathbf{1}_{\beta^T X \geq 0, \beta_n^T X < 0} \right). \end{aligned}$$

Because the second term on the right-hand side is clearly nonpositive, we have that $M_n(\beta) \leq M_n(\beta_n)$ for all $\beta \in \mathcal{S}^{d-1}$. Now let $\epsilon > 0$ and consider a compact set \mathbf{X}_ϵ such that $\mathbb{Q}_n(\mathbf{X}_\epsilon \times \mathbb{R}) > 1 - \epsilon$ for all $n \in \mathbb{N}$ (its existence is guaranteed by (A1)). Then, the identity $\mathbb{Q}_n(f_{\beta, \beta_n}) = \mathbb{Q}_n((\kappa_n(X) - 1/2)\mathbf{1}_{\beta^T X \geq 0})$ implies $|M_n(\beta) - \mathbb{Q}_n((\kappa(X) - 1/2)\mathbf{1}_{\beta^T X \geq 0})| \leq 2\mathbb{Q}_n((\mathbb{R}^d \setminus \mathbf{X}_\epsilon) \times \mathbb{R}) + \|\kappa_n - \kappa\|_{\mathbf{X}_\epsilon}$ for all $\beta \in \mathcal{S}^{d-1}$. Consequently, (A1), (A2) and (A3) put together show that $\overline{\lim} \|M_n - \mathbb{Q}_n((\kappa(X) - 1/2)\mathbf{1}_{(\cdot)^T X \geq 0})\|_{\mathcal{S}^{d-1}} \leq 2\epsilon$. But $\epsilon > 0$ was arbitrarily chosen and we also have $\|\mathbb{Q}_n - \mathbb{P}\|_{\mathcal{F}} \rightarrow 0$, so we therefore have $\|M_n - \Gamma\|_{\mathcal{S}^{d-1}} \rightarrow 0$. Considering that β_0 is the unique maximizer of the continuous function Γ we can conclude the desired result as β_n maximizes M_n and the argmax function is continuous on continuous functions with unique maximizers (under the sup-norm). \square

3.1 Consistency and rate of convergence

We now introduce some additional notation. Denote by \mathbb{P}_n^* the empirical measure defined by the row $(X_{n,1}, U_{n,1}) \dots, (X_{n,m_n}, U_{n,m_n})$. Note that a vector $\beta_n^* \in \mathcal{S}^{d-1}$ is a maximum score estimator of β_n if it satisfies

$$\beta_n^* = \operatorname{argmax}_{\beta \in \mathcal{S}^{d-1}} \{\mathbb{P}_n^*(f_{\beta, \beta_n})\}.$$

We are now in a position to state our first consistency result.

Lemma 3.2 *If $\{A1-A3\}$ hold, $\beta_n^* \xrightarrow{\mathbf{P}} \beta_0$.*

Proof: Consider the classes of functions $\mathcal{F}_1 := \{\mathbf{1}_{\beta^T x + u \geq 0, \alpha^T x \geq 0} : \alpha, \beta \in \mathbb{R}^d\}$ and $\mathcal{F}_2 := \{\mathbf{1}_{\alpha^T x \geq 0} : \alpha \in \mathbb{R}^d\}$. Since the class of all half-spaces of \mathbb{R}^{d+1} is VC (see Exercise 14, page 152 in [Van der Vaart and Wellner \(1996\)](#)), Lemma 2.6.18 in page 147 of [Van der Vaart and Wellner \(1996\)](#) implies that $\mathcal{F}_1 = \{\mathbf{1}_{\beta^T x + u \geq 0} : \beta \in \mathbb{R}^d\} \wedge \{\mathbf{1}_{\alpha^T x \geq 0} : \alpha \in \mathbb{R}^d\}$ and \mathcal{F}_2 are both VC-subgraph classes of functions. Since these classes have the constant one as measurable envelope, they are *Euclidian* and, thus, *manageable* in the sense Definition 7.9 in page 38 of [Pollard \(1990\)](#) and hence, the maximal inequality 7.10 in page 38 of [Pollard \(1990\)](#) implies the existence of two positive constants $J_1, J_2 < \infty$ such that $\mathbf{E}(\|\mathbb{P}_n^* - \mathbb{Q}_n\|_{\mathcal{F}_1}) \leq \frac{J_1}{\sqrt{m_n}}$ and $\mathbf{E}(\|\mathbb{P}_n^* - \mathbb{Q}_n\|_{\mathcal{F}_2}) \leq \frac{J_2}{\sqrt{m_n}}$. As $\mathcal{F} = \mathcal{F}_1 - \frac{1}{2}\mathcal{F}_2$ this implies that $\|\mathbb{P}_n^* - \mathbb{Q}_n\|_{\mathcal{F}} \xrightarrow{\mathbf{P}} 0$. Considering that $\|M_n - \Gamma\|_{\mathcal{S}^{d-1}} \rightarrow 0$ (as shown in the proof of Lemma 3.1), the result follows from an application of Corollary 3.2.3 in page 287 of [Van der Vaart and Wellner \(1996\)](#). \square

We will now deduce the rate of convergence of β_n^* . To this end, we introduce the functions $\Gamma_n : \mathbb{R}^d \rightarrow \mathbb{R}$ given by

$$\Gamma_n(\beta) := \mathbb{P}((\kappa_n(X) - 1/2) \mathbf{1}_{\beta^T X \geq 0}).$$

It will be shown that β_n^* converges at rate $m_n^{-1/3}$. The proof of this fact relies on empirical processes arguments like those used to prove Lemma 4.1 in [Kim and Pollard \(1990\)](#). To adapt those ideas to our context (a triangular array with independent,

non-identically distributed rows) we need a maximal inequality specially designed for this situation. This is given in the following Lemma (proved in Section A.1).

Lemma 3.3 *Under {A1-A5}, there is a constant $C_{R_0} > 0$ such that for any $R > 0$ and $n \in \mathbb{N}$ such that $\Delta_0 m_n^{-1/3} \leq R m_n^{-1/3} \leq R_0$ we have*

$$\mathbf{E} \left(\sup_{|\beta_n - \beta| \leq R m_n^{-1/3}} \{ |(\mathbb{P}_n^* - \mathbb{Q}_n)(f_{\beta, \beta_n} - f_{\beta_n, \beta_n})| \}^2 \right) \leq C_{R_0} R m_n^{-4/3} \quad \forall n \in \mathbb{N}.$$

With the aid of Lemma 3.3 we can now derive the rate of convergence of the maximum score estimator.

Lemma 3.4 *Under {A1-A5}, $m_n^{1/3}(\beta_n^* - \beta_n) = O_{\mathbf{P}}(1)$.*

Proof: Take R_0 as in (A4), let $\epsilon > 0$ and define

$$M_{\epsilon, n} := \inf \left\{ a > 0 : \sup_{\substack{|\beta - \beta_n| \leq R_0 \\ \beta \in \mathcal{S}^{d-1}}} \{ |(\mathbb{P}_n^* - \mathbb{Q}_n)(f_{\beta, \beta_n} - f_{\beta_n, \beta_n})| - \epsilon |\beta - \beta_n|^2 \} \leq a m_n^{-2/3} \right\};$$

$$B_{n, j} := \{ \beta \in \mathcal{S}^{d-1} : (j-1)m_n^{-1/3} < |\beta - \beta_n| \leq j m_n^{-1/3} \wedge R_0 \}.$$

Then, by Lemma 3.3 we have

$$\begin{aligned} & \mathbf{P}(M_{\epsilon, n} > a) \\ &= \mathbf{P}(\exists \beta \in \mathcal{S}^{d-1}, |\beta - \beta_n| \leq R_0 : |(\mathbb{P}_n^* - \mathbb{Q}_n)(f_{\beta, \beta_n} - f_{\beta_n, \beta_n})| > \epsilon |\beta - \beta_n|^2 + a^2 m_n^{-2/3}) \\ &\leq \sum_{j=1}^{\infty} \mathbf{P}(\exists \beta \in B_{n, j} : m_n^{2/3} |(\mathbb{P}_n^* - \mathbb{Q}_n)(f_{\beta, \beta_n} - f_{\beta_n, \beta_n})| > \epsilon^2 (j-1)^2 + a^2) \\ &\leq \sum_{j=1}^{\infty} \frac{m_n^{4/3}}{(\epsilon(j-1)^2 + a^2)^2} \mathbf{E} \left(\sup_{|\beta_n - \beta| < j m_n^{-1/3} \wedge R_0} \{ |(\mathbb{P}_n^* - \mathbb{Q}_n)(f_{\beta, \beta_n} - f_{\beta_n, \beta_n})| \}^2 \right) \\ &\leq C_{R_0} \sum_{j=1}^{\infty} \frac{j}{(\epsilon(j-1)^2 + a^2)^2} \rightarrow 0 \text{ as } a \rightarrow \infty. \end{aligned}$$

It follows that $M_{\epsilon, n} = O_{\mathbf{P}}(1)$. On the other hand, by (ii) and (iii) of (A4) and the fact that $\mathbb{Q}_n(f_{\beta, \beta_n}) = \mathbb{Q}_n((\kappa_n(X) - 1/2)\mathbf{1}_{\beta^T X \geq 0})$ imply that

$$|(\mathbb{Q}_n(f_{\beta, \beta_n} - f_{\beta_n, \beta_n}) - (\Gamma_n(\beta) - \Gamma_n(\beta_n)))| \leq \frac{3}{2} \epsilon_n m_n^{-1/3} |\beta - \beta_n|$$

for all $n \in \mathbb{N}$ and $\Delta_0 m_n^{-1/3} \leq |\beta - \beta_n| \leq R_0$. Also, Lemma A.2 implies that

$$|(\Gamma_n - \Gamma)(\beta_n^*) - (\Gamma_n - \Gamma)(\beta_n)| \leq |\beta_n^* - \beta_n|^{(\nu-1)/\nu} o(m_n^{-(\nu+1)/3\nu}).$$

On the other hand, Lemma A.1 implies that $\nabla^2 \Gamma$ is continuous around β_0 and that there is a convex, open set $U \subset \mathbb{R}^d$ containing β_0 such that $\nabla^2 \Gamma(\beta)$ is negative definite on $U \setminus \{t\beta_n : t \geq 0\}$. Without loss of generality, assume in addition $\epsilon > 0$ is small enough so that the largest eigenvalue of $\nabla^2 \Gamma(\beta)$ is smaller than -2ϵ for all $\beta \in U \setminus \{t\beta_n : t \geq 0\}$. Hence, as $\beta_n^* \xrightarrow{\mathbf{P}} \beta_0$ and $\beta_n \rightarrow \beta_0$, there is $N > 0$ such that for all $n \geq N$ we have that $\beta_n, \beta_n^* \in U \setminus \{t\beta_n : t \geq 0\}$ and

$$\Gamma_n(\beta_n^*) - \Gamma_n(\beta_n) < |\beta_n^* - \beta_n|^{(\nu-1)/\nu} o(m_n^{-(\nu+1)/3\nu}) + |\nabla \Gamma(\beta_n)| |\beta_n^* - \beta_n| - 2\epsilon |\beta_n^* - \beta_n|^2,$$

with probability tending to one. Putting all these facts together, and considering Lemma 3.2, we conclude that, with probability tending to one, whenever $|\beta_n^* - \beta_n| \geq \Delta_0 m_n^{-1/3}$, we also have

$$\begin{aligned} & \mathbb{P}_n^*(f_{\beta_n^*, \beta_n} - f_{\beta_n, \beta_n}) \\ & \leq \mathbb{Q}_n(f_{\beta_n^*, \beta_n} - f_{\beta_n, \beta_n}) + \epsilon |\beta_n^* - \beta_n|^2 + M_{\epsilon, n}^2 m_n^{-2/3} \\ & \leq \Gamma_n(\beta_n^*) - \Gamma_n(\beta_n) + \frac{3}{2} \epsilon_n |\beta_n^* - \beta_n| m_n^{-1/3} + \epsilon |\beta_n^* - \beta_n|^2 + M_{\epsilon, n}^2 m_n^{-2/3} \\ & \leq |\beta_n^* - \beta_n|^{(\nu-1)/\nu} o(m_n^{-(\nu+1)/3\nu}) + |\nabla \Gamma(\beta_n)| |\beta_n^* - \beta_n| - 2\epsilon |\beta_n^* - \beta_n|^2 \\ & \quad + \frac{3}{2} \epsilon_n |\beta_n^* - \beta_n| m_n^{-1/3} + \epsilon |\beta_n^* - \beta_n|^2 + M_{\epsilon, n}^2 m_n^{-2/3} \\ & \leq -\epsilon |\beta_n^* - \beta_n|^2 + \frac{3}{2} \epsilon_n |\beta_n^* - \beta_n| m_n^{-1/3} + M_{\epsilon, n}^2 m_n^{-2/3} + |\nabla \Gamma(\beta_n)| |\beta_n^* - \beta_n| \\ & \quad + \Delta_0^{-1/\nu} m_n^{-1/3} |\beta_n^* - \beta_n| O(1). \end{aligned}$$

Finally, using (A5) and the fact that $M_{\epsilon, n} = O_{\mathbf{P}}(1)$ and $\mathbb{P}_n^*(f_{\beta_n^*, \beta_n} - f_{\beta_n, \beta_n}) \geq 0$, we can conclude that with probability tending to one as $n \rightarrow \infty$ we can write

$$\epsilon |\beta_n^* - \beta_n|^2 \leq O_{\mathbf{P}}(m_n^{-1/3}) |\beta_n^* - \beta_n| + O_{\mathbf{P}}(m_n^{-2/3}).$$

The result now easily follows. \square

3.2 Asymptotic distribution

Before going into the derivation of the limit law of β_n^* , we need to introduce some further notation. Consider a sequence of matrices $(H_n)_{n=1}^\infty \subset \mathbb{R}^{d \times (d-1)}$ and $H \in \mathbb{R}^{d \times (d-1)}$ satisfying the following properties:

- (a) $\xi \mapsto H_n \xi$ and $\xi \mapsto H \xi$ are bijections from \mathbb{R}^{d-1} to the hyperplanes $\{x \in \mathbb{R}^d : \beta_n^T x = 0\}$ and $\{x \in \mathbb{R}^d : \beta_0^T x = 0\}$, respectively.
- (b) The columns of H_n and H form orthonormal bases for $\{x \in \mathbb{R}^d : \beta_n^T x = 0\}$ and $\{x \in \mathbb{R}^d : \beta_0^T x = 0\}$, respectively.
- (c) There is a constant $C_H > 0$, depending only on H , such that $\|H_n - H\|_2 \leq C_H |\beta_n - \beta_0|$.

We now give an intuitive argument for the existence of such a sequence of matrices. Imagine that for β_0 we find an orthonormal basis $\{e_{0,1}, \dots, e_{0,d-1}\}$ for the hyperplane $\{x \in \mathbb{R}^d : \beta_0^T x = 0\}$ and we let H have these vectors as columns. We then obtain the *rigid motion* $\mathcal{T} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ that moves β_0 to β_n and the hyperplane $\{x \in \mathbb{R}^d : \beta_0^T x = 0\}$ to $\{x \in \mathbb{R}^d : \beta_n^T x = 0\}$. We let the columns of H_n be given by $\{\mathcal{T}e_{0,1}, \dots, \mathcal{T}e_{0,d-1}\}$. The resulting sequence of matrices will satisfy the (a), (b) and (c) for some constant C_H .

Note that (b) implies that H_n^T and H^T are the Moore-Penrose pseudo-inverses of H_n and H , respectively. In particular, $H_n^T H_n = H^T H = \mathcal{I}_{d-1}$, where \mathcal{I}_{d-1} is the identity matrix in \mathbb{R}^{d-1} (in the sequel we will always use this notation for identity matrices on Euclidian spaces). Additionally, it can be inferred from (b) that $H_n^T (\mathcal{I}_d - \beta_n \beta_n^T) = H_n^T$ and $H^T (\mathcal{I}_d - \beta_0 \beta_0^T) = H^T$. Now, for each $s \in \mathbb{R}^{d-1}$ define $\beta_{n,s} := \left(\sqrt{1 - (m_n^{-1/3} |s|)^2} \wedge 1 \beta_n + m_n^{-1/3} H_n s \right) \mathbf{1}_{|s| \leq m_n^{1/3}} + |s|^{-1} H_n s \mathbf{1}_{|s| > m_n^{1/3}}$. Note that as long as $|s| \leq m_n^{1/3}$ we have $\beta_{n,s} \in \mathcal{S}^{d-1}$ with $H_n s$ being the orthogonal projection of $\beta_{n,s}$ onto the hyperplane generated by β_n . Define the process $\Lambda_n(s) := m_n^{2/3} \mathbb{P}_n^*(f_{\beta_{n,s}, \beta_n} - f_{\beta_n, \beta_n})$. Observe that if $(\beta_n^*)_{n=1}^\infty$ is a sequence of maximum

score estimators, then with probability tending to one as $n \rightarrow \infty$ we have

$$s_n^* := m_n^{1/3} H_n^T (\mathcal{I}_d - \beta_n \beta_n^T) (\beta_n^* - \beta_n) = m_n^{1/3} H_n^T (\beta_n^* - \beta_n) = \underset{s \in \mathbb{R}^{d-1}}{\operatorname{argmax}} \{\Lambda_n(s)\}. \quad (3)$$

Considering this, we will regard the processes Λ_n as random elements in the space of locally bounded real-valued functions on \mathbb{R}^{d-1} (denoted by $\mathbf{B}_{loc}(\mathbb{R}^{d-1})$) and then derive the limit law of s_n^* by applying the argmax continuous mapping theorem. We will take the space $\mathbf{B}_{loc}(\mathbb{R}^{d-1})$ with the topology of uniform convergence on compacta. Our approach is based on that in [Kim and Pollard \(1990\)](#).

To properly describe the asymptotic distribution we need to define the function $\Sigma : \mathbb{R}^{d-1} \times \mathbb{R}^{d-1} \rightarrow \mathbb{R}$ as follows:

$$\begin{aligned} \Sigma(s, t) &:= \frac{1}{4} \int_{\mathbb{R}^{d-1}} \{[(s^T \xi) \wedge (t^T \xi)]_+ + [(s^T \xi) \vee (t^T \xi)]_-\} p(H\xi) \, d\xi \\ &= \frac{1}{8} \int_{\mathbb{R}^{d-1}} (|s^T \xi| + |t^T \xi| - |(s-t)^T \xi|) p(H\xi) \, d\xi. \end{aligned}$$

Additionally, denote by W_n the $\mathbf{B}_{loc}(\mathbb{R}^{d-1})$ -valued process given by $W_n(s) := m_n^{2/3} (\mathbb{P}_n^* - \mathbb{Q}_n)(f_{\beta_n, s, \beta_n} - f_{\beta_n, \beta_n})$. In what follows, the symbol \rightsquigarrow will denote distributional convergence. We are now in a position to state and prove our convergence theorem.

Theorem 3.1 *Assume that {A1-A5} hold. Then, there is a $\mathbf{B}_{loc}(\mathbb{R}^{d-1})$ -valued stochastic process Λ of the form $\Lambda(s) = W(s) + \frac{1}{2} s^T H^T \nabla^2 \Gamma(\beta_0) H s$, where W is a zero-mean Gaussian process in $\mathbf{B}_{loc}(\mathbb{R}^{d-1})$ with continuous sample paths and covariance function Σ . Moreover, Λ has a unique maximizer with probability one and we have*

- (i) $\Lambda_n \rightsquigarrow \Lambda$ in $\mathbf{B}_{loc}(\mathbb{R}^{d-1})$.
- (ii) $s_n^* \rightsquigarrow s^* := \underset{s \in \mathbb{R}^{d-1}}{\operatorname{argmax}} \{\Lambda(s)\}$.
- (iii) $m_n^{1/3} (\beta_n^* - \beta_n) \rightsquigarrow H s^*$.

Proof: Lemmas [A.5](#) and [A.6](#) imply that the sequence $(W_n)_{n=1}^\infty$ is stochastically equicontinuous and that its finite dimensional distributions converge to those of a zero-mean Gaussian process with covariance Σ . From Theorem 2.3 in [Kim and Pollard \(1990\)](#) we know that here exists a continuous process W with these properties

and such that $W_n \rightsquigarrow W$. Moreover, from Lemma A.3 (i) and (iii) we can easily deduce that $\Lambda_n - W_n - m_n^{2/3}(\Gamma_n(\beta_{n,(\cdot)}) - \Gamma_n(\beta_n)) \xrightarrow{\mathbf{P}} 0$ and $m_n^{2/3}(\Gamma_n(\beta_{n,(\cdot)}) - \Gamma_n(\beta_n)) \xrightarrow{\mathbf{P}} \frac{1}{2}(\cdot)^T H^T \nabla^2 \Gamma(\beta_0) H(\cdot)$ on $\mathbf{B}_{loc}(\mathbb{R}^{d-1})$ (with the topology of convergence on compacta). Thus, applying Slutsky's Lemma (see Example 1.4.7, page 32 in [Van der Vaart and Wellner \(1996\)](#)) we get that $\Lambda_n \rightsquigarrow \Lambda$. The uniqueness of the maximizers of the sample paths of Λ follows from Lemmas 2.5 and 2.6 in [Kim and Pollard \(1990\)](#). Finally, an application of Theorem 2.7 in [Kim and Pollard \(1990\)](#) gives (ii); (iii) follows from (3). \square

As a corollary we immediately get the asymptotic distribution of the maximum score estimator (by taking $\kappa_n = \kappa$ and $\beta_n = \beta_0$).

Corollary 3.1 *If $(X_n, U_n)_{n=1}^\infty \stackrel{i.i.d.}{\sim} \mathbb{P}$ and $(\hat{\beta}_n)_{n=1}^\infty$ is a sequence of maximum score estimators computed from $(X_n, Y_n)_{n=1}^\infty$ then, $n^{1/3}(\hat{\beta}_n - \beta_0) \rightsquigarrow H \operatorname{argmax}_{s \in \mathbb{R}^{d-1}} \{\Lambda(s)\}$.*

One final remark is to be made about the process Λ . The quadratic drift term can be rewritten, by using the matrix H to evaluate the surface integral, to obtain the following more convenient expression

$$\Lambda(s) = W(s) - \frac{1}{2} s^T \left(\int_{\mathbb{R}^{d-1}} (\nabla \kappa(H\xi)^T \beta_0) p(H\xi) \xi \xi^T d\xi \right) s.$$

4 Consistency of the bootstrap procedures

In this section we study the consistency of the bootstrap procedures proposed in Section 2.2. We recall the notation and definitions established in Section 2.2. We will show that schemes 2, 3 and 4 in Section 2.2 are consistent. The classical bootstrap scheme is known to be *inconsistent* for the maximum score estimator. This is argued in [Abrevaya and Huang \(2005\)](#). The subsampling scheme for this problem (see Section 2.2.5) has been analyzed in [Delgado et al. \(2001\)](#). Consistency of the m -out-of- n bootstrap for the maximum score estimator can be deduced from the results in [Lee and Pun \(2006\)](#).

4.1 Fixed design residual bootstrap

We will prove the consistency of this bootstrap scheme appealing to Theorem 3.1. For the proof of the following result, recall the notation and definitions in Section 2.2.2.

Theorem 4.1 *Under (A0) and the conditions of the fixed design residual bootstrap (see Section 2.2.2), the conditional distribution of $n^{1/3}(\beta_n^* - \hat{\beta}_n)$ given $(X_1, Y_1), \dots, (X_n, Y_n)$ consistently estimates the distribution of $n^{1/3}(\hat{\beta}_n - \beta_0)$.*

Proof: Set $\mathbb{Q}_{n,j}$ to be a probability measure in \mathbb{R}^{d+1} such that whenever $(X, U) \sim \mathbb{Q}_{n,j}$ we have that $X = X_j$ (that is, X has as distribution the dirac measure concentrated at X_j) and $\mathbb{Q}_{n,j}(\hat{\beta}_n^T X + U \geq 0 | X) = \hat{\kappa}_n(X_j)$. By the remark made at the end of Section 2.2.2, the bootstrap sample can be seen as a realization of independent random vectors of the form $(X_{n,j}^*, \mathbf{1}_{\hat{\beta}_n^T X_{n,j}^* + U_{n,j}^* \geq 0})$ with $(X_{n,j}^*, U_{n,j}^*) \sim \mathbb{Q}_{n,j}$. Note that in this case the marginal of X under \mathbb{Q}_n is $\mathbb{Q}_n((\cdot) \times \mathbb{R}) = \mu_n(\cdot)$. Lemmas B.2 and B.3, together with the properties of $\tilde{\kappa}_n$ in (i) of Section 2.2.2, imply that every subsequence $(\mathbb{Q}_{n_k})_{k=1}^\infty$ has a further subsequence $(\mathbb{Q}_{n_{k_s}})_{s=1}^\infty$ for which properties {A1-A4} in Section 3 hold with probability one (here one takes $\nu \in (2, r)$ with $r > 2$ as in (A0) and sets $m_s = n_{k_s}$, $\beta_s = \hat{\beta}_{n_{k_s}}$, $\mathbb{Q}_{s,j} = \mathbb{Q}_{n_{k_s},j}$ and $\kappa_s = \hat{\kappa}_{n_{k_s}}$). To complete the argument, we would like to use Lemma B.1 to show that every subsequence $(\mathbb{Q}_{n_k})_{k=1}^\infty$ has a further subsequence for which (A5) holds almost surely. However, Lemma B.1 only states that $\nabla\Gamma(\hat{\beta}_n) = O_{\mathbf{P}}(1)$ and this fact alone does not allow us to prove such a claim. To remedy this situation, further justification is required and we give it in the following paragraph.

Let $\delta, \eta > 0$ and $A_n = [|\nabla\Gamma(\hat{\beta}_n)| > \eta]$. Choose η large enough so that $\mathbf{P}(A_n) < \delta$ for all $n \in \mathbb{N}$. Note that each A_n is measurable with respect to the σ -algebra generated by $(X_1, Y_1), \dots, (X_n, Y_n)$. Denote by ρ the Prokhorov metric and by λ_n and λ_n^* , respectively, the law of $n^{1/3}(\hat{\beta}_n - \beta_0)$ and the conditional distribution of $n^{1/3}(\beta_n^* - \hat{\beta}_n)$ given the data. Finally, define $\tilde{\lambda}_n := \lambda_n \mathbf{1}_{A_n} + \lambda_n^* \mathbf{1}_{\Omega \setminus A_n}$. For each $n \in \mathbb{N}$, $\tilde{\lambda}_n$ can be thought as the conditional law (given the data) of $n^{1/3}(\tilde{\beta}_n - \hat{\beta}_n \mathbf{1}_{\Omega \setminus A_n} - \beta_0 \mathbf{1}_{A_n})$, where $\tilde{\beta}_n$ is the

maximum score estimator from the sample $(\tilde{X}_{n,1}, \tilde{Y}_{n,1}), \dots, (\tilde{X}_{n,n}, \tilde{Y}_{n,n})$ defined to be $(X_1, Y_1), \dots, (X_n, Y_n)$ should event A_n happen and $(X_{n,1}^*, Y_{n,1}^*), \dots, (X_{n,n}^*, Y_{n,n}^*)$ otherwise. Then, the definition of A_n and the arguments in the preceding paragraph imply that for every subsequence of the sequence of laws generating $(\tilde{X}_{n,1}, \tilde{Y}_{n,1}), \dots, (\tilde{X}_{n,n}, \tilde{Y}_{n,n})$ there is a further subsequence for which {A1-A5} hold almost surely. Hence, an application of Theorem 3.1 shows that $\rho(\tilde{\lambda}_n, \lambda_n) \xrightarrow{\mathbf{P}} 0$. But then, $\mathbf{P}(\rho(\lambda_n^*, \lambda_n) > \delta) \leq \mathbf{P}(A_n) + \mathbf{P}(\rho(\tilde{\lambda}_n, \lambda_n) > \delta)$ for all $n \in \mathbb{N}$ and hence $\overline{\lim} \mathbf{P}(\rho(\lambda_n^*, \lambda_n) > \delta) \leq \delta$. As $\delta > 0$ was arbitrarily chosen, this finishes the proof. \square

4.2 Random design residual bootstrap

The consistency of this procedure is given in the following result.

Theorem 4.2 *Under (A0) and the conditions of the random design residual bootstrap (see Section 2.2.2), the conditional distribution of $n^{1/3}(\beta_n^* - \hat{\beta}_n)$ given $(X_1, Y_1), \dots, (X_n, Y_n)$ consistently estimates the distribution of $n^{1/3}(\hat{\beta}_n - \beta_0)$.*

Proof: Let \mathbb{Q}_n be the probability measure on \mathbb{R}^{d+1} such that if $(X, U) \sim \mathbb{Q}_n$ then $X \sim \mu_n$ and $\mathbf{E}\left(\mathbf{1}_{\hat{\beta}_n^T X + U \geq 0} \mid X = X_j\right) = \hat{\kappa}_n(X_j)$ for all $j = 1, \dots, n$. Then, by the remark at the end of Section 2.2.2, we can regard the bootstrap sample $(X_{n,1}^*, Y_{n,1}^*), \dots, (X_{n,n}^*, Y_{n,n}^*)$ as a realization of the form $(X_{n,1}^*, \mathbf{1}_{\hat{\beta}_n^T X_{n,1}^* + U_{n,1}^* \geq 0}), \dots, (X_{n,n}^*, \mathbf{1}_{\hat{\beta}_n^T X_{n,n}^* + U_{n,n}^* \geq 0})$ with $(X_{n,1}^*, U_{n,1}^*), \dots, (X_{n,n}^*, U_{n,n}^*) \stackrel{i.i.d.}{\sim} \mathbb{Q}_n$. Considering this, we can apply Theorem 3.1 in a similar fashion as in the proof of Theorem 4.2 to complete the argument. \square

4.3 The smooth bootstrap

As in the previous section, we will use Theorem 3.1. Consider the notation of Section 2.2.4. Note that the remark made regarding the condition that $\hat{\beta}_n^T x(\hat{\kappa}_n(x) - 1/2) \geq 0$ in 2.2.2 also applies for the smooth bootstrap. Let \mathbb{Q}_n be the probability measure on

\mathbb{R}^{d+1} such that if $(X, U) \sim \mathbb{Q}_n$ then $X \sim \hat{F}_n$ and $\mathbf{E} \left(\mathbf{1}_{\hat{\beta}_n^T X + U \geq 0} \middle| X \right) = \hat{\kappa}_n(X)$. Then, we can regard the bootstrap samples as $(X_{n,1}^*, \mathbf{1}_{\hat{\beta}_n^T X_{n,1}^* + U_{n,1}^* \geq 0}), \dots, (X_{n,n}^*, \mathbf{1}_{\hat{\beta}_n^T X_{n,n}^* + U_{n,n}^* \geq 0})$ with $(X_{n,1}^*, U_{n,1}^*), \dots, (X_{n,n}^*, U_{n,n}^*) \stackrel{i.i.d.}{\sim} \mathbb{Q}_n$. For this scheme we have the following result.

Theorem 4.3 *Consider the conditions of the smooth bootstrap scheme (see Section 2.2.4) and assume that (A0) holds. Then, the conditional distribution of $n^{1/3}(\beta_n^* - \hat{\beta}_n)$ given $(X_1, Y_1), \dots, (X_n, Y_n)$ consistently estimates the distribution of $n^{1/3}(\hat{\beta}_n - \beta_0)$.*

Proof: By setting $\kappa_n = \hat{\kappa}_n$, $\beta_n = \hat{\beta}_n$, $\nu \in (2, r)$ and $\mathbb{Q}_{n,j} = \mathbb{Q}_n$ one can see with the aid of Lemma B.2 that every sequence of these objects has a further subsequence that satisfies {A1-A3} with probability one. As for (A4), it is easily seen that for any function f which is uniformly bounded by a constant K we have

$$|(\mathbb{Q}_n - \mathbb{P})(f(X))| = \left| \int_{\mathbb{R}^d} f(x)(\hat{p}_n(x) - p(x)) dx \right| \leq K \|\hat{p}_n - p\|_1.$$

It is now straightforward to show that (A4) will hold in probability because $\|\hat{p}_n - p\|_1 = O_{\mathbf{P}}(\varepsilon_n)$. The proof can then be completed by arguing as in the last paragraph of the proof of Theorem 4.1. \square

5 Simulation Experiments

In this section we illustrate the finite sample performance of the bootstrap methods introduced in Section 2.2 through simulation experiments. We take the distribution \mathbb{P} on \mathbb{R}^3 (so $d = 2$ in this case) to satisfy the assumptions of our model with $\beta_0 = \frac{1}{\sqrt{2}}(1, 1)^T$ and such that if $(X, U) \sim \mathbb{P}$ then $U|X \sim N\left(0, \frac{1}{(1+|X|^2)^2}\right)$ and $X \sim \text{Uniform}([-1, 1]^2)$. Thus, in this case $\kappa(x) = 1 - \Phi(-\beta_0^T x(1 + |x|^2))$ which is, of course, infinitely differentiable. Consequently, according to Stone (1982), the optimal (achievable) rates of convergence to estimate κ nonparametrically are faster than those required in (i) of Section 2.2.2. To compute the estimator $\tilde{\kappa}_n$ of κ we have

chosen to use the Nadaraya-Watson estimator with a Gaussian kernel and a bandwidth given by Scott's normal reference rule (see [Scott \(1992\)](#), page 152). We would like to point out that our selection of the smoothing bandwidth could be improved by applying data-driven selection methods, such as cross-validation.

We start by showing a plot of the 75%-quantile of the conditional distribution of the first component of $n^{1/3}(\beta_n^* - \hat{\beta}_n)$ under the different bootstrap schemes and compare it to that of $n^{1/3}(\hat{\beta}_n - \beta_0)$ for 5 different sample paths as n increases from 2,000 to 30,000, when sampling is done from the distribution \mathbb{P} . [Figure 1](#) shows the 5 plots, each containing 5 piecewise linear curves: the black (diamond-shape dots), blue (circular dots) and red (triangular dots) dotted curves correspond to the classical bootstrap, the random and the fixed design residual bootstraps (schemes 2 and 3), respectively. The green (square dots) dotted line and (broken) purple line correspond, respectively, to the 75%-quantile of the distribution of the first coordinate of $n^{1/3}(\hat{\beta}_n - \beta_0)$ and its asymptotic limit. From each of the conditional distributions 1,000 bootstrap samples were used to produce the empirical estimates of the corresponding quantile. The erratic behavior of the classical bootstrap is clearly illustrated in the plots, and this provides numerical evidence of the fact that the classical bootstrap has no weak limit in probability. On the other hand, the fixed-design and random design residual bootstraps exhibit pretty stable paths. The curves corresponding to the quantiles produced by these two schemes are fluctuating around the green and broken lines with quite reasonable deviations. This is an indication of their convergence. It must be noted that the fact that $\hat{\beta}_n$ exhibits cube-root asymptotics implies that one must go to considerably large sample sizes to obtain reasonable accuracy.

In addition to [Figure 1](#), we will now provide another graphical tool that will illustrate the convergence of the different bootstrap schemes. We took a sample of size $n = 2000$ from \mathbb{P} and constructed histograms from the 1000 bootstrap samples for 7 different schemes: the classical bootstrap, the random design and fixed design residual bootstrap methods, the smooth bootstrap and m-out-of-n bootstrap with $m_n = \lceil \sqrt{n} \rceil$, $\lceil n^{2/3} \rceil$, $\lceil n^{4/5} \rceil$. As before, we used the Nadaraya-Watson estimator with

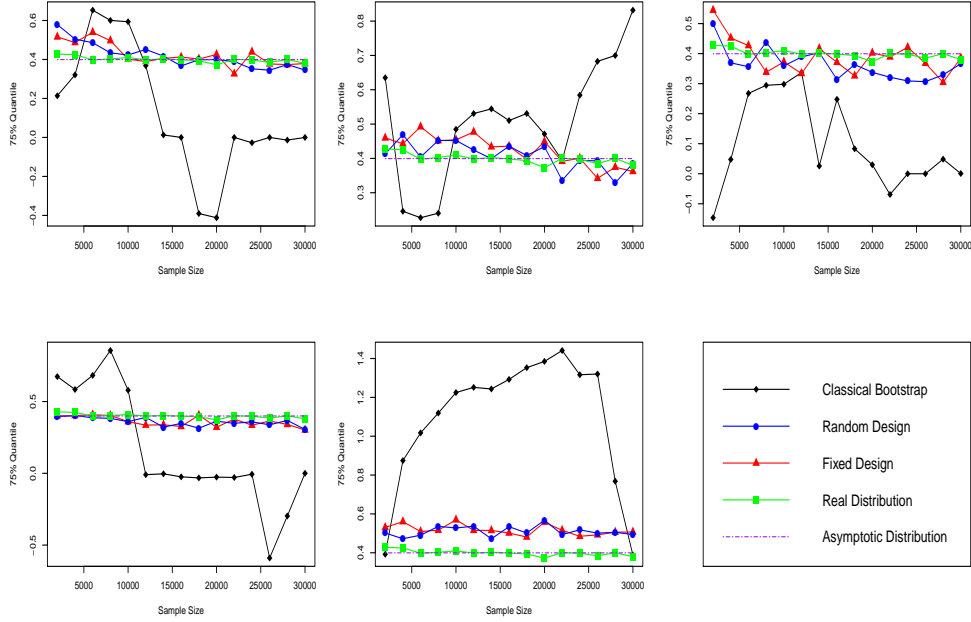


Figure 1: Plots of the 75%-quantiles for the conditional distributions of $n^{1/3}(\beta_{n,1}^* - \hat{\beta}_{n,1})$ for 5 different sample paths as n increases from 2,000 to 30,000 (by increments of 2,000). The piecewise linear curves correspond to the classical bootstrap (black), the fixed design bootstrap (red), the random design bootstrap, together with the corresponding quantiles of $n^{1/3}(\hat{\beta}_{n,1} - \beta_{0,1})$ (green) and of the asymptotic distribution (purple broken line).

Scott's normal reference rule and Gaussian kernels to build $\tilde{\kappa}_n$ in the cases of random and design schemes and the smooth bootstrap. For the density estimation step in the smooth bootstrap, we used a kernel density estimator with Gaussian kernels and Scott's normal reference rule for the bandwidth. Similarly, we chose the different values of m_n randomly. Both, the selection of bandwidth and of m_n could have been improved by using data-driven procedures. In particular, the use of crossed-validation could enhance the performance of the bandwidth-dependent methods. In addition to all this, the corresponding graphics were also built for 1000 random samples of the first component of $n^{1/3}(\hat{\beta}_n - \beta_0)$, with $n = 2000$, and its asymptotic distribution. In the case of the \mathbb{P} described above, the asymptotic distribution of the first component of $n^{1/3}(\hat{\beta}_n - \beta_0)$ is that of $\frac{1}{\sqrt{2}} \operatorname{argmax}_{s \in \mathbb{R}} \{\Lambda(s)\}$ with $\Lambda(s) := 2^{-5/4} Z(s) - \frac{11}{30\sqrt{\pi}} s^2$,

where Z is a standard two-sided Brownian motion. The resulting histograms are displayed in Figure 2.

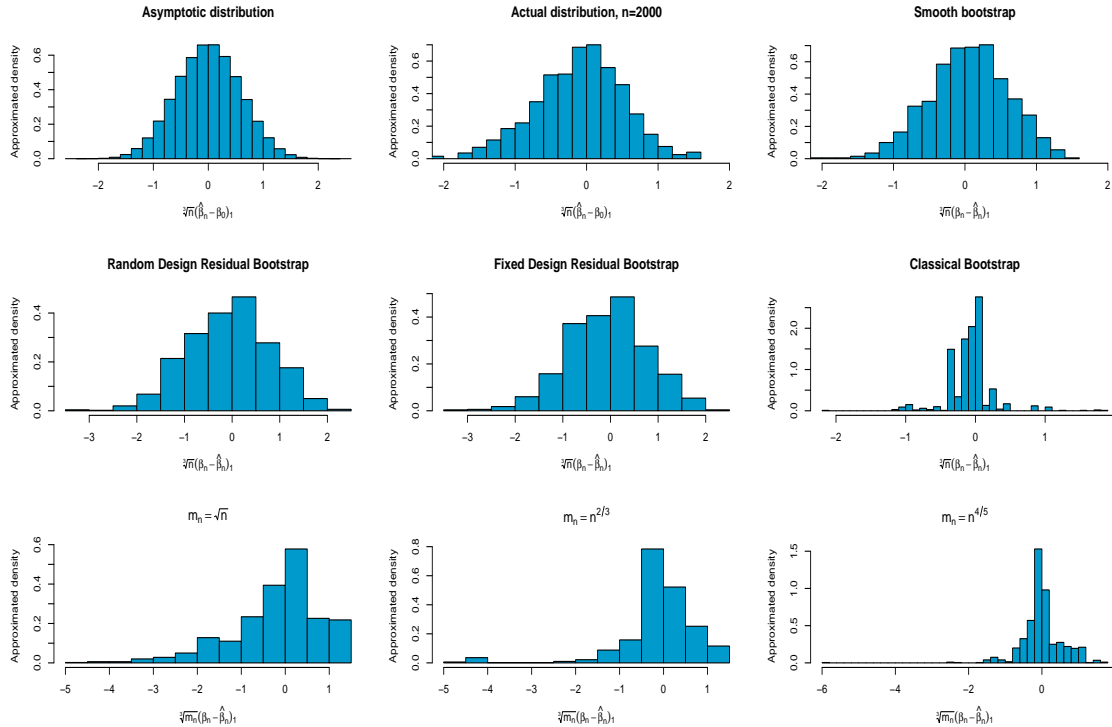


Figure 2: Histograms for the conditional distributions of $n^{1/3}(\beta_{n,1}^* - \hat{\beta}_{n,1})$ for random samples of size 1000 with bootstrap distributions built from a sample of size $n = 2000$ from \mathbb{P} . Histograms of random samples from $n^{1/3}(\hat{\beta}_{n,1} - \beta_{0,1})$ and its asymptotic distribution are also included.

It is clear from Figure 2 that the histogram from the smooth bootstrap (top-right) is the one that best approximates those from both, the actual distribution of $n^{1/3}(\hat{\beta}_{n,1} - \beta_{0,1})$, $n = 2000$ (top-center) and its asymptotic distribution (top-left). The graphic corresponding to the former seems to provide the best approximation among the different bootstrap schemes not only to the shape, but also to the range of the latter two. Figure 2 also makes immediately apparent the lack of convergence of the classical bootstrap as its histogram (middle-right) is quite different from the ones in the top row. The random design (middle-left) and the fixed design (middle-center) residual bootstrap schemes give reasonable approximations to the shape of the histograms in the top row, albeit their x -range is slightly greater. Although known to

converge, the m-out-of-n schemes (bottom row) give visibly asymmetric histograms with large range, compared to the other convergent schemes. This fact will translate in generally larger, more conservative confidence intervals (at least for large sample sizes).

We now study the performance of each of the bootstrap schemes by measuring the average length and coverage of confidence intervals built from several random samples from \mathbb{P} as above. We simulated 1000 different random samples of sizes 100, 200, 500, 1000, 2000 and 5000. For each of these samples 7 different confidence intervals were built: one using the classical bootstrap, one using the random design scheme, one via the fixed design procedure, another one with the smooth bootstrap scheme and three for the m-out-of-n bootstrap with $m_n = \lceil n^{1/2} \rceil, \lceil n^{2/3} \rceil, \lceil n^{4/5} \rceil$. We have made no attempt to choose an optimal m_n . The interested reader can look at [Delgado et al. \(2001\)](#) for some data-driven procedures to choose m_n (although in a subsampling context). We would like to add that, as before, we have used the Nadaraya-Watson estimator with Scott's normal reference rule and Gaussian kernels to compute $\tilde{\kappa}_n$ in Section 2.2. In the particular case of the smooth bootstrap, we estimated the underlying density using a kernel density estimator with Gaussian kernel and Scott's normal rule. As before, our bandwidth and kernel selection methods could be improved by using data-driven procedures. In addition to considering a \mathbb{P} as the one used above, we have conducted the same experiments with one in which $U|X \sim (1 + |X|^2)^{-1}\Xi$, $\Xi \sim Student(3)$, $X \sim Uniform([-1, 1]^2)$, $\beta_0 = 2^{-1/2}(1, 1)^T$, where *Student*(3) stands for a standard Student-*t* distribution with 3 degrees of freedom. The results are reported in Table 1.

Table 1 indicates that the smooth bootstrap scheme outperforms all the others as it achieves the best combination of high coverage and small average length. At small sample sizes, in both cases, its coverage is similar to those of the random and the fixed design bootstraps and higher than those of the subsampling schemes. Its average length is, overall, considerably smaller than those of all the other consistent

Table 1: The estimated coverage probabilities and average lengths of nominal 95% CIs for the first coordinate of β_0 obtained using the four different bootstrap schemes for each of the two models.

$U X \sim N(0, (1 + X ^2)^{-2}), X \sim Uniform([-1, 1]^2), \beta_0 = 2^{-1/2}(1, 1)^T$						
Scheme	$n = 100$		$n = 200$		$n = 500$	
	Coverage	Avg Length	Coverage	Avg Length	Coverage	Avg Length
Classical	0.68	0.91	0.75	0.58	0.75	0.49
Random Design	0.80	1.09	0.90	0.85	0.91	0.72
Fixed Design	0.80	1.06	0.89	0.83	0.90	0.69
Smooth	0.79	0.57	0.89	0.44	0.93	0.30
$\lceil n^{1/2} \rceil$	0.73	0.93	0.83	0.72	0.87	0.58
$\lceil n^{2/3} \rceil$	0.72	0.89	0.83	0.70	0.87	0.56
$\lceil n^{4/5} \rceil$	0.70	0.87	0.84	0.70	0.85	0.51
$U X \sim N(0, (1 + X ^2)^{-2}), X \sim Uniform([-1, 1]^2), \beta_0 = 2^{-1/2}(1, 1)^T$						
Scheme	$n = 1000$		$n = 2000$		$n = 5000$	
	Coverage	Avg Length	Coverage	Avg Length	Coverage	Avg Length
Classical	0.73	0.26	0.71	0.19	0.71	0.13
Random Design	0.95	0.43	0.93	0.29	0.94	0.19
Fixed Design	0.94	0.41	0.93	0.29	0.94	0.19
Smooth	0.95	0.23	0.94	0.18	0.95	0.13
$\lceil n^{1/2} \rceil$	0.91	0.41	0.89	0.33	0.97	0.23
$\lceil n^{2/3} \rceil$	0.95	0.46	0.92	0.36	0.95	0.21
$\lceil n^{4/5} \rceil$	0.89	0.34	0.86	0.24	0.89	0.16
$U X \sim (1 + X ^2)^{-1}\Xi, \Xi \sim Student(3), X \sim Uniform([-1, 1]^2), \beta_0 = 2^{-1/2}(1, 1)^T$						
Scheme	$n = 100$		$n = 200$		$n = 500$	
	Coverage	Avg Length	Coverage	Avg Length	Coverage	Avg Length
Classical	0.66	0.82	0.70	0.73	0.74	0.41
Random Design	0.78	1.17	0.85	0.90	0.93	0.65
Fixed Design	0.78	1.14	0.85	0.88	0.92	.62
Smooth	0.78	0.57	0.89	0.43	0.93	0.29
$\lceil n^{1/2} \rceil$	0.74	0.98	0.77	0.78	0.86	0.59
$\lceil n^{2/3} \rceil$	0.73	0.96	0.77	0.74	0.87	0.58
$\lceil n^{4/5} \rceil$	0.72	0.92	0.79	0.74	0.90	0.62
$U X \sim (1 + X ^2)^{-1}\Xi, \Xi \sim Student(3), X \sim Uniform([-1, 1]^2), \beta_0 = 2^{-1/2}(1, 1)^T$						
Scheme	$n = 1000$		$n = 2000$		$n = 5000$	
	Coverage	Avg Length	Coverage	Avg Length	Coverage	Avg Length
Classical	0.72	0.34	0.74	0.21	0.70	0.14
Random Design	0.94	0.55	0.94	0.33	0.94	0.21
Fixed Design	0.93	0.52	0.93	0.31	0.94	0.21
Smooth	0.93	0.23	0.94	0.18	0.95	0.13
$\lceil n^{1/2} \rceil$	0.87	0.50	0.92	0.36	0.93	0.27
$\lceil n^{2/3} \rceil$	0.92	0.55	0.94	0.40	0.97	0.26
$\lceil n^{4/5} \rceil$	0.88	0.49	0.88	0.29	0.87	0.17

procedures. However, in spite of its superior performance, the smooth bootstrap has the drawback of its computational complexity. As the sample size and dimension increase, the smooth bootstrap is considerably more complicated to implement than the rest.

The random and the fixed design bootstraps yield similar results. The difference between them being that random design bootstrap is a little bit more conservative: it achieves slightly better coverage at the expense of larger confidence intervals. This difference seems to disappear as the sample size increases. To compare these procedures with the m-out-of-n schemes, we first contrast their performance for the three smallest sample sizes and then for the three largest: for the small sample sizes ($n = 100, 200, 500$), the random and fixed design procedures are more conservative (better coverage with larger intervals) than the m-out-of-n schemes; as the sample size increases, the situation is reversed: the intervals from the former two achieve the asymptotic 95% coverage with generally less average length. Needless to say, the classical bootstrap performs poorly compared to the others.

An obvious conclusion of the previous analysis is that the smooth bootstrap is the best choice whenever it is computationally feasible. Compared to the m-out-of-n schemes, the random and fixed design procedures have the advantage of seemingly achieving the desired coverage faster and having smaller intervals for larger sample sizes. In the absence of a good data-driven rule to select m_n , the former two seem like a better option than the latter. Moreover, the fixed design procedure could be a better alternative than its random counterpart as it is computationally less expensive to implement (it requires less simulations).

Remark: Although we did not choose m_n using any specific rule, we did try many possible choices besides those listed in Table 1. More precisely, we also tried $m_n = \lceil n^{1/3} \rceil$, $\lceil n^{9/10} \rceil$ and $\lceil n^{14/15} \rceil$ but omitted a report of their results as they were all outperformed by the reported cases ($\lceil n^{1/2} \rceil$, $\lceil n^{2/3} \rceil$ and $\lceil n^{4/5} \rceil$).

A Auxiliary results for the proof of Theorem 3.1

Lemma A.1 Consider the functions $\Gamma(\beta) = \mathbb{P}(f_{\beta, \beta_0})$ and $\Gamma_n(\beta) = \mathbb{P}\left(\left(\kappa_n(X) - \frac{1}{2}\right) \mathbf{1}_{\beta^T X \geq 0}\right)$. Denote by σ_β the surface measure on the hyperplane $\{x \in \mathbb{R}^d : \beta^T x = 0\}$. For each $\alpha, \beta \in \mathbb{R}^d \setminus \{0\}$ define the matrix $A_{\alpha, \beta} := (\mathcal{I}_d - |\beta|^{-2} \beta \beta^T)(\mathcal{I}_d - |\alpha|^{-2} \alpha \alpha^T) + |\beta|^{-1} |\alpha|^{-1} \beta \alpha^T$. Then, Γ is twice continuously differentiable on \mathcal{S}^{d-1} , β_0 is the only maximizer of Γ on \mathcal{S}^{d-1} and we have

$$\begin{aligned} \nabla \Gamma(\beta) &= \frac{\beta^T \beta_0}{|\beta|^2} \left(\mathcal{I}_d - \frac{1}{|\beta|^2} \beta \beta^T \right) \int_{\beta_0^T x = 0} \left(\kappa(A_{\beta_0, \beta} x) - \frac{1}{2} \right) p(A_{\beta_0, \beta} x) x \, d\sigma_{\beta_0}; \\ &= \frac{\beta^T \beta_0}{|\beta|^2} \left(\mathcal{I}_d - \frac{1}{|\beta|^2} \beta \beta^T \right) \int_{\mathbb{R}^{d-1}} \left(\kappa(A_{\beta_0, \beta} H \xi) - \frac{1}{2} \right) p(A_{\beta_0, \beta} H \xi) H \xi \, d\xi \quad \forall \beta \in \mathcal{S}^{d-1}; \\ \nabla^2 \Gamma(\beta_0) &= - \int_{\beta_0^T x = 0} (\nabla \kappa(x)^T \beta_0) p(x) x x^T \, d\sigma_{\beta_0} \\ &= - \int_{\mathbb{R}^{d-1}} (\nabla \kappa(H \xi)^T \beta_0) p(H \xi) H \xi \xi^T H^T \, d\xi; \end{aligned}$$

where $H \in \mathbb{R}^{d \times d-1}$ is any matrix whose columns form an orthonormal basis of the hyperplane $\{x \in \mathbb{R}^d : \beta_0^T x = 0\}$. Furthermore, there is an open neighborhood $U \subset \mathbb{R}^d$ of β_0 such that $\beta^T \nabla^2 \Gamma_0(\beta_0) \beta < 0$ for all $\beta \in U \setminus \{t\beta_0 : t \in \mathbb{R}\}$.

Proof: Lemma 2 in [Manski \(1985\)](#) implies that β_0 is the only minimizer of Γ on \mathcal{S}^{d-1} . The differentiability properties of Γ as well as the computations of $\nabla \Gamma$ and $\nabla^2 \Gamma$ follow from the arguments in Section 5, page 205 and Example 6.4, page 213 in [Kim and Pollard \(1990\)](#). Note that for any x with $\beta_0^T x = 0$ we have $\nabla \kappa(x)^T \beta_0 \geq 0$ (because for x orthogonal to β_0 , $\kappa(x + t\beta_0) \leq 1/2$ and $\kappa(x + t\beta_0) \geq 1/2$ whenever $t < 0$ and $t > 0$, respectively). Additionally, for any $\beta \in \mathbb{R}^d$ we have:

$$\beta^T \nabla^2 \Gamma(\beta_0) \beta = - \int_{\beta_0^T x = 0} (\nabla \kappa(x)^T \beta_0) (\beta^T x)^2 p(x) \, d\sigma_{\beta_0}.$$

Thus, the fact that the set $\{x \in \mathfrak{X}^\circ : (\nabla \kappa(x)^T \beta_0) p(x) > 0\}$ is open (as p and $\nabla \kappa$ are continuous) and intersects the hyperplane $\{x \in \mathbb{R}^d : \beta_0^T x = 0\}$ implies that $\nabla^2 \Gamma(\beta_0)$ is negative definite on a set of the form $U \setminus \{t\beta_0 : t \in \mathbb{R}\}$ with $U \subset \mathbb{R}^d$ being an open neighborhood of β_0 . \square

Lemma A.2 Consider the functions Γ_n defined in Section 3.1. Then, under {A1-A5}

(i) $\beta_n = \operatorname{argmax}_{\beta \in \mathcal{S}^{d-1}} \{\Gamma_n(\beta)\};$

(ii) There is R_0 such that for all $0 < R \leq R_0$ we have

$$\sup_{|\beta - \beta_n| \leq R} \{ |(\Gamma_n - \Gamma)(\beta) - (\Gamma_n - \Gamma)(\beta_n)| \} \leq R^{\frac{\nu-1}{\nu}} o\left(m_n^{-\frac{1+\nu}{3\nu}}\right).$$

Proof: Note that we can write

$$\begin{aligned} \Gamma_n(\beta) &= \mathbb{P}\left(\left(\mathbb{Q}_{n,1}(U \geq -\beta_n^T X|X) - \mathbb{Q}_{n,1}(U \geq 0|X)\right)\mathbf{1}_{\beta^T X \geq 0, \beta_n^T X \geq 0}\right) + \\ &\quad \mathbb{P}\left(\left(\mathbb{Q}_{n,1}(U \geq -\beta_n^T X|X) - \mathbb{Q}_{n,1}(U \geq 0|X)\right)\mathbf{1}_{\beta^T X \geq 0, \beta_n^T X < 0}\right). \end{aligned}$$

It clearly follows that β_n is a maximizer of Γ_n . On the other hand, (A3), our assumptions on \mathbb{P} and Hölder's inequality imply that there are constants $C, R_0 > 0$ such that whenever $R \leq R_0$ we have

$$\sup_{|\beta - \beta_n| \leq R} \{ |(\Gamma_n - \Gamma)(\beta) - (\Gamma_n - \Gamma)(\beta_n)| \} \leq \mathbb{P}(|\kappa_n(X) - \kappa(X)|^\nu)^{1/\nu} \mathbb{P}(F_{n,R})^{\frac{\nu-1}{\nu}} \leq (CR)^{\frac{\nu-1}{\nu}} o\left(m_n^{-\frac{\nu+1}{3\nu}}\right).$$

This finishes the proof. □

Lemma A.3 Let $R > 0$ and consider the notation of Section 3.2. Then, under {A1-A5} we have

(i) $\sup_{|s| \leq R} \left\{ m_n^{2/3} \left| \left(\mathbb{Q}_n - \mathbb{P} \right) \left(\left(\kappa_n(X) - \frac{1}{2} \right) \left(\mathbf{1}_{\beta_{n,s}^T X \geq 0} - \mathbf{1}_{\beta_n^T X \geq 0} \right) \right) \right| \right\} \rightarrow 0;$

(ii) $\sup_{|s| \leq R} \left\{ m_n^{2/3} \left| \mathbb{Q}_n \left(\left(\kappa_n(X) - \frac{1}{2} \right) \left(\mathbf{1}_{\beta_{n,s}^T X \geq 0} - \mathbf{1}_{\beta_n^T X \geq 0} \right) \right) - \mathbb{P} \left(\left(\kappa(X) - \frac{1}{2} \right) \left(\mathbf{1}_{\beta_{n,s}^T X \geq 0} - \mathbf{1}_{\beta_n^T X \geq 0} \right) \right) \right| \right\} \rightarrow 0.$

(iii) $\sup_{|s| \leq R} \left\{ m_n^{2/3} \left| \Gamma(\beta_{n,s}) - \Gamma(\beta_n) - \frac{1}{2} s^T H_n^T \nabla^2 \Gamma(\beta_0) H_n s \right| \right\} \rightarrow 0.$

Proof: Observe that $|\beta_{n,s} - \beta_n|^2 = |s|^2 m_n^{-2/3} + \left(\sqrt{1 - (m_n^{-1/3}|s|)^2} \wedge 1 - 1 \right)^2$ for all n and s . It follows that $\sup_{|s| \leq R} \{ |\beta_{n,s} - \beta_n| \} = O(R m_n^{-1/3})$. Hence, in view of (A4), there is a constant $K > 0$ such that for all n large enough

$$\sup_{|s| \leq R} \left\{ m_n^{2/3} \left| \left(\mathbb{Q}_n - \mathbb{P} \right) \left(\left(\kappa_n(X) - 1/2 \right) \left(\mathbf{1}_{\beta_{n,s}^T X \geq 0} - \mathbf{1}_{\beta_n^T X \geq 0} \right) \right) \right| \right\} \leq \epsilon_n(KR) \vee \Delta_n \rightarrow 0.$$

Since (ii) is a direct consequence of Lemma A.2, it only remains to prove (iii). From Taylor's theorem we know that for every $s \in \mathbb{R}^d$ there is $\theta_{n,s} \in [0, 1]$ such that

$\Gamma(\beta_{n,s}) = \Gamma(\beta_n) + \nabla\Gamma(\beta_n)^T(\beta_{n,s} - \beta_n) + \frac{1}{2}(\beta_{n,s} - \beta_n)^T \nabla^2\Gamma(\theta_{n,s}\beta_{n,s} + (1 - \theta_{n,s})\beta_n)(\beta_{n,s} - \beta_n)$. It follows that

$$\begin{aligned} & \sup_{|s| \leq R} \left\{ m_n^{2/3} \left| \Gamma(\beta_{n,s}) - \Gamma(\beta_n) - \frac{1}{2}(\beta_{n,s} - \beta_n)^T \nabla^2\Gamma(\beta_0)(\beta_{n,s} - \beta_n) \right| \right\} \\ & \leq \sup_{|s| \leq R} \{ |\beta_{n,s} - \beta_n| |\nabla\Gamma(\beta_n)| \} + \sup_{|s| \leq R} \{ |\beta_{n,s} - \beta_n|^2 \} \sup_{|s| \leq R} \{ \|\nabla^2\Gamma(\beta_{n,s}) - \nabla^2\Gamma(\beta_0)\|_2 \} \rightarrow 0, \end{aligned}$$

where the convergence to 0 follows from (A5) and the fact that $\Gamma(\cdot)$ is twice continuously differentiable in a neighborhood of β_0 . But from the definition of $\beta_{n,s}$ it is easily seen that

$$\sup_{|s| \leq R} \left\{ m_n^{2/3} \left| \frac{1}{2} s^T H_n^T \nabla^2\Gamma(\beta_0) H_n s - \frac{1}{2}(\beta_{n,s} - \beta_n)^T \nabla^2\Gamma(\beta_0)(\beta_{n,s} - \beta_n) \right| \right\} \rightarrow 0.$$

This finishes the argument. \square

Lemma A.4 *Let $R > 0$. Under $\{A1-A5\}$ there is a sequence of random variables $\Delta_n^R = O_{\mathbf{P}}(1)$ such that for every $\delta > 0$ and every $n \in \mathbb{N}$ we have,*

$$\sup_{\substack{|s-t| \leq \delta \\ |s| \vee |t| \leq R}} \{ \mathbb{P}_n^* ((f_{\beta_{n,s},\beta_n} - f_{\beta_{n,t},\beta_n})^2) \} \leq \delta \Delta_n^R m_n^{-1/3}.$$

Proof: Define $\mathcal{G}_{R,\delta}^n := \{f_{\beta_{n,s},\beta_n} - f_{\beta_{n,t},\beta_n} : |s-t| \leq \delta, |s| \vee |t| \leq R\}$ and $\mathcal{G}_R^n := \{f_{\beta_{n,s},\beta_n} - f_{\beta_{n,t},\beta_n} : |s| \vee |t| \leq R\}$. It can be shown that \mathcal{G}_R^n is manageable with envelope $G_{n,R} := 3F_{n,2Rm_n^{-1/3}}$ (as $|k_n - 1/2| \leq 1$). Note that $G_{n,R}$ is independent of δ . Then (A4) and the maximal inequality 7.10 from Pollard (1990) then there is a constant \tilde{J}_R such that for all large enough n we have

$$\begin{aligned} \mathbf{E} \left(\sup_{\substack{|s-t| \leq \delta \\ |s| \vee |t| \leq R}} \{ \mathbb{P}_n^* ((f_{\beta_{n,s},\beta_n} - f_{\beta_{n,t},\beta_n})^2) \} \right) & \leq 2\mathbf{E} \left(\sup_{\substack{|s-t| \leq \delta \\ |s| \vee |t| \leq R}} \{ \mathbb{P}_n^* (|f_{\beta_{n,s},\beta_n} - f_{\beta_{n,t},\beta_n}|) \} \right); \\ & \leq 2\mathbf{E} \left(\sup_{f \in \mathcal{G}_R^n} \{ (\mathbb{P}_n^* - \mathbb{Q}_n) |f| \} \right) + 2 \sup_{f \in \mathcal{G}_{R,\delta}^n} \{ \mathbb{Q}_n |f| \}; \\ & \leq \frac{4\epsilon_1 \tilde{J}_R \sqrt{(\epsilon_1 + C) R m_n^{-1/3}}}{\sqrt{m_n}} + 2\epsilon_n R m_n^{-1/3} + 2 \sup_{f \in \mathcal{G}_{R,\delta}^n} \{ \mathbb{P} |f| \}. \end{aligned}$$

On the other hand, our assumptions on \mathbb{P} imply that the function $\mathbb{P}(\mathbf{1}_{(\cdot)^T x \geq 0})$ is continuously differentiable, and hence Lipschitz, on \mathcal{S}^{d-1} . Thus, there is a constant

L , independent of δ , such that

$$\mathbf{E} \left(\sup_{\substack{|s-t| \leq \delta \\ |s| \vee |t| \leq R}} \left\{ \mathbb{P}_n^* \left((f_{\beta_n, s, \beta_n} - f_{\beta_n, t, \beta_n})^2 \right) \right\} \right) \leq o(m_n^{-1/3}) + \delta L m_n^{-1/3}.$$

The result now follows. \square

Lemma A.5 *Under $\{A1-A5\}$, for every $R, \epsilon, \eta > 0$ there is $\delta > 0$ such that*

$$\overline{\lim}_{n \rightarrow \infty} \mathbf{P} \left(\sup_{\substack{|s-t| \leq \delta \\ |s| \vee |t| \leq R}} \left\{ m_n^{2/3} |(\mathbb{P}_n^* - \mathbb{Q}_n)(f_{\beta_n, s, \beta_n} - f_{\beta_n, t, \beta_n})| \right\} > \eta \right) \leq \epsilon.$$

Proof: Let $\Psi_n := m_n^{1/3} \mathbb{P}_n^*(4F_{n, Rm_n^{-1/3}}^2) = m_n^{1/3} \mathbb{P}_n^*(F_{n, Rm_n^{-1/3}})$. Note that our assumptions on \mathbb{P} then imply that there is a constant \tilde{C} such that $\mathbb{P}(F_{n, K}^2) = \mathbb{P}(F_{n, K}) \leq \tilde{C}CK$ for $0 < K \leq K_*$ ($F_{n, K}$ is an indicator function). Considering this, assumption (A4)-(i) and Lemma 3.3 imply that

$$\mathbf{E}(\Psi_n) = m_n^{1/3} \mathbf{E} \left((\mathbb{P}_n^* - \mathbb{Q}_n)(F_{n, Rm_n^{-1/3}}) \right) + m_n^{1/3} (\mathbb{Q}_n - \mathbb{P})(F_{n, Rm_n^{-1/3}}) + m_n^{1/3} \mathbb{P}(F_{n, Rm_n^{-1/3}}) = O(1).$$

Now, define $\Phi_n := m_n^{1/3} \sup_{\substack{|s-t| \leq \delta \\ |s| \vee |t| \leq R}} \left\{ \mathbb{P}_n^* \left((f_{\beta_n, s, \beta_n} - f_{\beta_n, t, \beta_n})^2 \right) \right\}$. The class of all differences $f_{\beta_n, s, \beta_n} - f_{\beta_n, t, \beta_n}$ with $|s| \vee |t| \leq R$ and $|s - t| < \delta$ is manageable (in the sense of definition 7.9 in page 38 of Pollard (1990)) for the envelope function $2F_{n, Rm_n^{-1/3}}$. By the maximal inequality 7.10 in Pollard (1990), there is a continuous increasing function J with $J(0) = 0$ and $J(1) < \infty$ such that

$$\mathbf{E} \left(\sup_{\substack{|s-t| \leq \delta \\ |s| \vee |t| \leq R}} \left\{ |(\mathbb{P}_n^* - \mathbb{Q}_n)(f_{\beta_n, s, \beta_n} - f_{\beta_n, t, \beta_n})| \right\} \right) \leq \frac{1}{m_n^{2/3}} \mathbf{E} \left(\sqrt{\Psi_n} J(\Phi_n / \Psi_n) \right).$$

Let $\rho > 0$. Breaking the integral on the right on the events that $\Psi_n \leq \rho$ and $\Psi_n > \rho$ and the applying Cauchy-Schwartz inequality,

$$\begin{aligned} \mathbf{E} \left(\sup_{\substack{|s-t| \leq \delta \\ |s| \vee |t| \leq R}} \left\{ m_n^{2/3} |(\mathbb{P}_n^* - \mathbb{Q}_n)(f_{\beta_n, s, \beta_n} - f_{\beta_n, t, \beta_n})| \right\} \right) &\leq \sqrt{\rho} J(1) + \sqrt{\mathbf{E}(\Psi_n \mathbf{1}_{\Psi_n > \rho})} \sqrt{\mathbf{E}(J(1 \wedge (\Phi_n / \rho)))}, \\ &\leq \sqrt{\rho} J(1) + \sqrt{\mathbf{E}(\Psi_n \mathbf{1}_{\Psi_n > \rho})} \sqrt{\mathbf{E}(J(1 \wedge (\delta \Delta_n^R / \rho)))}, \end{aligned}$$

where $\Delta_n^R = O_{\mathbf{P}}(1)$ is as in Lemma A.4. It follows that for any given $R, \eta, \epsilon > 0$ we can choose ρ and δ small enough so that the results holds. \square

Lemma A.6 *Let $s, t, s_1, \dots, s_N \in \mathbb{R}^{d-1}$ and write $\Sigma_N \in \mathbb{R}^{N \times N}$ for the matrix given by $\Sigma_N := (\Sigma(s_k, s_j))_{k,j}$. Then, under $\{A1-A5\}$ we have*

$$(a) \ m_n^{1/3} \mathbb{Q}_n(f_{\beta_n, s, \beta_n} - f_{\beta_n, \beta_n}) \rightarrow 0,$$

$$(b) \ m_n^{1/3} \mathbb{Q}_n((f_{\beta_n, s, \beta_n} - f_{\beta_n, \beta_n})(f_{\beta_n, t, \beta_n} - f_{\beta_n, \beta_n})) \rightarrow \Sigma(s, t),$$

$$(c) \ (W_n(s_1), \dots, W_n(s_N))^T \rightsquigarrow N(0, \Sigma_N),$$

where $N(0, \Sigma_N)$ denotes an \mathbb{R}^N -valued Gaussian random vector with mean 0 and covariance matrix Σ_N and \rightsquigarrow stands for weak convergence.

Proof: (a) From Lemma A.2 (ii) it is easily seen that

$$m_n^{1/3} \mathbb{Q}_n(f_{\beta_n, s, \beta_n} - f_{\beta_n, \beta_n}) = m_n^{1/3} \mathbb{P} \left(\left(\kappa(X) - \frac{1}{2} \right) (\mathbf{1}_{\beta_n^T X \geq 0} - \mathbf{1}_{\beta_n^T X \geq 0}) \right) + o(1).$$

We will therefore focus on deriving a limit for

$$m_n^{1/3} \mathbb{P} \left(\left(\kappa(X) - \frac{1}{2} \right) (\mathbf{1}_{\beta_n^T X \geq 0} - \mathbf{1}_{\beta_n^T X \geq 0}) \right).$$

Consider the transformations $T_n : \mathbb{R}^d \rightarrow \mathbb{R}^d$ given by $T_n(x) := (H_n^T x; \beta_n^T x)$, where $H_n^T x \in \mathbb{R}^{d-1}$ and $\beta_n^T x \in \mathbb{R}$. Note that T_n is an orthogonal transformation so $\det(T_n) = \pm 1$ and for any $\xi \in \mathbb{R}^{d-1}$ and $\eta \in \mathbb{R}$ we have $T_n^{-1}(\xi; \eta) = H_n \xi + \eta \beta_n$. Applying change of variables and Fubini's theorem, for all n large enough,

$$\begin{aligned} m_n^{1/3} \mathbb{P} \left(\left(\kappa(X) - \frac{1}{2} \right) (\mathbf{1}_{\beta_n^T X \geq 0} - \mathbf{1}_{\beta_n^T X \geq 0}) \right) &= \int_{\mathbb{R}^d} \left(\kappa(x) - \frac{1}{2} \right) (\mathbf{1}_{\beta_n^T x \geq 0} - \mathbf{1}_{\beta_n^T x \geq 0}) p(x) dx \\ &= m_n^{1/3} \iint (\kappa(H_n \xi + \eta \beta_n) - 1/2) (\mathbf{1}_{\sqrt{1-m_n^{-2/3}} |s|^2 \eta + m_n^{-1/3} s^T \xi \geq 0} - \mathbf{1}_{\eta \geq 0}) p(H_n \xi + \eta \beta_n) d\eta d\xi \\ &= \int_{\mathbb{R}^{d-1}} m_n^{1/3} \int \left(\kappa(H_n \xi + \eta \beta_n) - \frac{1}{2} \right) p(H_n \xi + \eta \beta_n) \left(\mathbf{1}_{\frac{-m_n^{-1/3} s^T \xi}{\sqrt{1-m_n^{-2/3}} |s|^2} \leq \eta < 0} + \mathbf{1}_{0 \leq \eta < \frac{-m_n^{-1/3} s^T \xi}{\sqrt{1-m_n^{-2/3}} |s|^2}} \right) d\eta d\xi \end{aligned}$$

The dominated convergence theorem then implies,

$$m_n^{1/3} \mathbb{P} \left(\left(\kappa(X) - \frac{1}{2} \right) (\mathbf{1}_{\beta_n^T X \geq 0} - \mathbf{1}_{\beta_n^T X \geq 0}) \right) \rightarrow - \int_{\mathbb{R}^{d-1}} \left(\kappa(H \xi) - \frac{1}{2} \right) p(H \xi) |s^T \xi| d\xi = 0,$$

where the last identity stems from the fact that $\kappa(x) := \mathbb{P}(U - \beta_0^T X \geq 0 | X = x)$ is identically $1/2$ on the hyperplane orthogonal to β_0 .

(b) First note that $(\mathbf{1}_{U+\beta_n^T X \geq 0} - 1/2)^2 \equiv 1/4$ and

$$(\mathbf{1}_{\beta_n^T s x \geq 0} - \mathbf{1}_{\beta_n^T t x \geq 0})(\mathbf{1}_{\beta_n^T t x \geq 0} - \mathbf{1}_{\beta_n^T x \geq 0}) = \mathbf{1}_{(\beta_n^T s x) \wedge (\beta_n^T t x) \geq 0 > \beta_n^T x} + \mathbf{1}_{\beta_n^T x \geq 0 > (\beta_n^T s x) \vee (\beta_n^T t x)}.$$

In view of these facts, condition (A4)-(iii) and the same change of variables as in the proof of (a) imply

$$\begin{aligned} & m_n^{1/3} \mathbb{Q}_n \left((f_{\beta_n, s, \beta_n} - f_{\beta_n, \beta_n})(f_{\beta_n, t, \beta_n} - f_{\beta_n, \beta_n}) \right) \\ &= \frac{m_n^{1/3}}{4} \mathbb{P}(\mathbf{1}_{(\beta_n^T s x) \wedge (\beta_n^T t x) \geq 0 > \beta_n^T x} + \mathbf{1}_{\beta_n^T x \geq 0 > (\beta_n^T s x) \vee (\beta_n^T t x)}) + o(1) \\ &= \frac{m_n^{1/3}}{4} \iint \left(\mathbf{1}_{-m_n^{-1/3} \frac{s^T \xi}{\sqrt{1-m_n^{-2/3}|s|^2}} \wedge \frac{t^T \xi}{\sqrt{1-m_n^{-2/3}|t|^2}} \leq \eta < 0} + \mathbf{1}_{0 \leq \eta < -m_n^{-1/3} \frac{s^T \xi}{\sqrt{1-m_n^{-2/3}|s|^2}} \vee \frac{t^T \xi}{\sqrt{1-m_n^{-2/3}|t|^2}}} \right) p(H_n \xi + \eta \beta_n) d\eta d\xi \\ & \quad + o(1). \end{aligned}$$

A further application of the dominated convergence theorem now yields

$$m_n^{1/3} \mathbb{Q}_n \left((f_{\beta_n, s, \beta_n} - f_{\beta_n, \beta_n})(f_{\beta_n, t, \beta_n} - f_{\beta_n, \beta_n}) \right) \rightarrow \frac{1}{4} \int_{\mathbb{R}^{d-1}} \left((s^T \xi \wedge t^T \xi)_+ + (s^T \xi \vee t^T \xi)_- \right) p(H \xi) d\xi$$

(c) Define $\zeta_n := (W_n(s_1), \dots, W_n(s_N))^T$; $\tilde{\zeta}_{n,k}$ to be the $(d-1)$ -dimensional random vector whose j -entry is $m_n^{-1/3}(f_{\beta_n, s_j, \beta_n}(X_{n,k}, U_{n,k}) - f_{\beta_n, \beta_n}(X_{n,k}, U_{n,k}))$; $\zeta_{n,k} := \tilde{\zeta}_{n,k} - \mathbf{E}(\tilde{\zeta}_{n,k})$; and

$$\rho_{n,k,j} := \mathbb{Q}_n \left((f_{\beta_n, s_k, \beta_n} - f_{\beta_n, \beta_n})(f_{\beta_n, s_j, \beta_n} - f_{\beta_n, \beta_n}) \right) - \mathbb{Q}_n \left((f_{\beta_n, s_k, \beta_n} - f_{\beta_n, \beta_n}) \right) \mathbb{Q}_n \left((f_{\beta_n, s_j, \beta_n} - f_{\beta_n, \beta_n}) \right).$$

We therefore have $\zeta_n = \sum_{k=1}^{m_n} \zeta_{n,k}$ and $\mathbf{E}(\zeta_{n,k}) = 0$. Moreover, (a) and (b) imply that $\sum_{k=1}^{m_n} \mathbf{Var}(\zeta_{n,k}) = \sum_{k=1}^{m_n} \mathbf{E}(\zeta_{n,k} \zeta_{n,k}^T) \rightarrow \Sigma_N$. Now, take $\theta \in \mathbb{R}^N$ and define $\alpha_{n,k} := \theta^T \zeta_{n,k}$. In the sequel we will denote by $\|\cdot\|_\infty$ the \mathbb{L}_∞ -norm on \mathbb{R}^N . The previous arguments imply that $\mathbf{E}(\alpha_{n,k}) = 0$ and that $s_n^2 := \sum_{k=1}^{m_n} \mathbf{Var}(\alpha_{n,k}) = \sum_{k=1}^{m_n} \theta^T \mathbf{Var}(\zeta_{n,k}) \theta \rightarrow \theta^T \Sigma_N \theta$. Finally, note that for all positive ϵ ,

$$\begin{aligned} \frac{1}{s_n} \sum_{l=1}^{m_n} \mathbf{E}(\alpha_{n,k}^2 \mathbf{1}_{|\alpha_{n,k}| > \epsilon s_n}) &\leq \frac{N^2 \|\theta\|_\infty^2 m_n^{-2/3}}{s_n} \sum_{l=1}^{m_n} \mathbb{Q}_{n,l}(|\alpha_{n,l}| > \epsilon s_n) \\ &\leq \frac{N^2 \|\theta\|_\infty^2 m_n^{-2/3}}{s_n^3 \epsilon^2} \sum_{1 \leq k, j \leq N} \theta_k \theta_j m_n^{1/3} \rho_{n,k,j} \rightarrow 0. \end{aligned}$$

By the Lindeberg-Feller central limit theorem we can thus conclude that $\theta^T \zeta_n = \sum_{j=1}^{m_n} \alpha_{n,j} \rightsquigarrow \mathbb{N}(0, \theta^T \Sigma_N \theta)$. Since $\theta \in \mathbb{R}^N$ was arbitrarily chosen, we can apply the

Cramer-Wold device to conclude (c). \square

A.1 Proof of Lemma 3.3

Let us take the notation of (A4). Take $R_0 \leq K_*$, so for any $K \leq R_0$ the class $\{f_{\beta, \beta_n} - f_{\beta_n, \beta_n}\}_{|\beta - \beta_n| < K}$ is majorized by $F_{n, K}$. Our assumptions on \mathbb{P} then imply that there is a constant \tilde{C} such that $\mathbb{P}(F_{n, K}^2) = \mathbb{P}(F_{n, K}) \leq \tilde{C}CK$ for $0 < K \leq K_*$ ($F_{n, K}$ is an indicator function). Now, take $R > 0$ and $n \in \mathbb{N}$ such that $\Delta_0 m_n^{-1/3} < R m_n^{-1/3} \leq R_0$. Since $F_{n, R m_n^{-1/3}, \Delta_0}$ is a VC-class (with VC index bounded by a constant independent of n and R), the maximal inequality 7.10 in page 38 of Pollard (1990) implies the existence of a constant J , not depending neither on m_n nor on R , such that $\mathbf{E} \left(\|\mathbb{P}_n^* - \mathbb{Q}_n\|_{\mathcal{F}_{n, R m_n^{-1/3}}}^2 \right) \leq J \mathbb{Q}_n(F_{n, R m_n^{-1/3}}) / m_n$. From (A4) we can conclude that $\mathbf{E} \left(\|\mathbb{P}_n^* - \mathbb{Q}_n\|_{\mathcal{F}_{n, R m_n^{-1/3}}}^2 \right) \leq J(O(m_n^{-1/3}) + \tilde{C}C R m_n^{-1/3}) / m_n$ for all R and n for which $m_n^{-1/3} R \leq R_0$. This finishes the proof. \square

B Auxiliary Results for the proof of the consistency of the bootstrap schemes

We present a simple result that will allow us to show that condition (A5) in Section 3 is satisfied for the bootstrap samples in Schemes 2, 3 and 4.

Lemma B.1 *Consider the function Γ defined in Section 3 and let $(\hat{\beta}_n)_{n=1}^\infty$ be a sequence of maximum score estimators. Then $|\nabla \Gamma(\hat{\beta}_n)| = O_{\mathbf{P}}(n^{-1/3})$.*

Proof: This follows from Lemma A.1 as $\nabla \Gamma(\beta_0) = 0$ and $\nabla \Gamma$ is continuously differentiable over the compact set \mathcal{S}^{d-1} (which implies that it is Lipschitz on that region).

\square

B.1 Properties of the modified regression estimator

Consider the estimators $\hat{\kappa}_n$ and $\tilde{\kappa}_n$ defined in Section 2.2.2. We will show in the following lemma that $\hat{\kappa}_n$ satisfies the same regularity conditions as $\tilde{\kappa}_n$.

Lemma B.2 *Consider the assumptions of section 2.2.2. Then we have,*

(i) *On a set with probability one, $\|\hat{\kappa}_n - \kappa\|_X \rightarrow 0$ on all compact sets $X \subset \mathfrak{X}^\circ$.*

(ii) $\mathbb{P}(|\hat{\kappa}_n - \kappa|^r)^{1/r} = O_{\mathbf{P}}(n^{-(r+1)/3r})$.

(iii) $\mu_n(|\hat{\kappa}_n - \kappa|^r)^{1/r} = O_{\mathbf{P}}(n^{-(r+1)/3r})$

Proof: Let $\delta > 0$ and $X \subset \mathfrak{X}^\circ$ be compact. Note that for all $x \in X$ we have

$$\begin{aligned} |\hat{\kappa}_n(x) - \kappa(x)| &\leq |\tilde{\kappa}_n(x) - \kappa(x)| + |\kappa(x) - 1/2| \mathbf{1}_{|\hat{\beta}_n^T x(\tilde{\kappa}_n(x)-1/2)| \leq \delta} + \\ &|\kappa(x) - 1/2| \mathbf{1}_{|\hat{\beta}_n^T x(\tilde{\kappa}_n(x)-1/2) - \beta_0^T x(\kappa(x)-1/2)| > \delta} + |\kappa(x) - 1/2| \mathbf{1}_{\beta_0^T x(\kappa(x)-1/2) \leq \delta}, \end{aligned}$$

from which it follows that

$$\begin{aligned} \|\hat{\kappa}_n - \kappa\|_X &\leq \|\tilde{\kappa}_n - \kappa\|_X + \mathbf{1}_{\inf_{x \in X} \{\hat{\beta}_n^T x(\tilde{\kappa}_n(x)-1/2)\} \leq \delta} + \\ &\mathbf{1}_{\sup_{x \in X} \{|\hat{\beta}_n^T x(\tilde{\kappa}_n(x)-1/2) - \beta_0^T x(\kappa(x)-1/2)|\} > \delta} + \sup_{\substack{\beta_0^T x(\kappa(x)-1/2) \leq \delta \\ x \in X}} \{|\kappa(x) - 1/2|\}. \end{aligned}$$

Thus, (i) is a consequence of the almost sure uniform convergence of $\tilde{\kappa}_n$ on compact subsets of \mathfrak{X}° .

We now turn our attention to (ii). It suffices to show that $I_n := \mathbb{P}(|\kappa(x) - 1/2|^r \mathbf{1}_{\hat{\beta}_n^T x(\tilde{\kappa}_n(x)-1/2) < 0})^{1/r} = O_{\mathbf{P}}(n^{-(r+1)/3r})$. Note that $|\beta_0^T x| |(\kappa(x)-1/2)| = \beta_0^T x(\kappa(x)-1/2)$. Hence, for any $\Delta > 0$ we have,

$$\begin{aligned} I_n &\leq \mathbb{P}(|\kappa(X) - 1/2|^r \mathbf{1}_{|\beta_0^T X| |(\kappa(X)-1/2)| \leq |\hat{\beta}_n^T X(\tilde{\kappa}_n(X)-1/2) - \beta_0^T X(\kappa(X)-1/2)|})^{1/r} \\ &\leq \mathbb{P}(|\kappa(X) - 1/2|^r \mathbf{1}_{|\beta_0^T X| |(\kappa(X)-1/2)| \leq |\hat{\beta}_n^T X(\tilde{\kappa}_n(X)-1/2) - \beta_0^T X(\kappa(X)-1/2)|, |\beta_0^T X| \leq \Delta n^{-(1+r)/3r}})^{1/r} + \\ &\quad \mathbb{P}(|\kappa(X) - 1/2|^r \mathbf{1}_{|\beta_0^T X| |(\kappa(X)-1/2)| \leq |\hat{\beta}_n^T X(\tilde{\kappa}_n(X)-1/2) - \beta_0^T X(\kappa(X)-1/2)|, |\beta_0^T X| > \Delta n^{-\frac{1+r}{3r}}})^{1/r} \\ &\quad \leq \mathbb{P}(|\kappa(X) - 1/2|^r \mathbf{1}_{|\beta_0^T X| \leq \Delta n^{-(1+r)/3r}})^{1/r} \\ + \mathbb{P}(|\kappa(X) - 1/2|^r \mathbf{1}_{|\beta_0^T X| |(\kappa(X)-1/2)| \leq |\hat{\beta}_n - \beta_0| \|X\| |\kappa(X) - 1/2| + |\hat{\beta}_n - \beta_0| \|X\| |\tilde{\kappa}_n(X) - \kappa(X)| + |\beta_0^T X| |\kappa(X) - \tilde{\kappa}_n(X)|, |\beta_0^T X| > \Delta n^{-\frac{1+r}{3r}}})^{1/r} \\ &\quad \leq \Delta n^{-(1+r)/3r} \mathbb{P}(D(X)^r \mathbf{1}_{|\beta_0^T X| \leq \Delta n^{-(1+r)/3r}})^{1/r} + \\ &\quad \mathbb{P}(|\kappa(X) - 1/2|^r \mathbf{1}_{|\beta_0^T X| \leq 3|\hat{\beta}_n - \beta_0| \|X\|})^{1/r} + \\ &\quad \mathbb{P}\left(|\kappa(X) - 1/2| \mathbf{1}_{|\kappa(X)-1/2| \leq 3\Delta^{-1} n^{-\frac{1+r}{3r}} |\hat{\beta}_n - \beta_0| \|X\| |\tilde{\kappa}_n(X) - \kappa(X)|}\right)^{1/r} + 3\mathbb{P}(|\tilde{\kappa}_n(X) - \kappa(X)|^r)^{1/r} \end{aligned}$$

$$\begin{aligned}
&\leq \Delta n^{-\frac{1+r}{3r}} \mathbb{P}(D(X)^r \mathbf{1}_{|\beta_0^T X| \leq \Delta n^{-(1+r)/3r}})^{1/r} + \\
&\quad \mathbb{P}(|\kappa(X) - 1/2|^r \mathbf{1}_{|\beta_0^T X| \leq 3|\hat{\beta}_n - \beta_0||X|})^{1/r} + \\
&3\Delta^{-1} |\hat{\beta}_n - \beta_0| n^{\frac{1+r}{3r}} \mathbb{P}(|X|^r \tilde{q})^{\frac{1}{r\tilde{q}}} \mathbb{P}(|\tilde{\kappa}_n(X) - \kappa(X)|^{r\tilde{q}})^{\frac{1}{r\tilde{q}}} + 3\mathbb{P}(|\tilde{\kappa}_n(X) - \kappa(X)|^r)^{\frac{1}{r}},
\end{aligned}$$

where $\tilde{q} = q/(q-1)$. The last inequality implies that

$$\begin{aligned}
I_n &\leq \Delta n^{-\frac{1+r}{3r}} \mathbb{P}(D(X)^r \mathbf{1}_{|\beta_0^T X| \leq \Delta n^{-(1+r)/3r}})^{1/r} + \\
&3|\hat{\beta}_n - \beta_0| \mathbb{P}(|X|^r D(X)^r \mathbf{1}_{|\beta_0^T X| \leq |\hat{\beta}_n - \beta_0||X|})^{1/r} + \\
&3\Delta^{-1} |\hat{\beta}_n - \beta_0| n^{\frac{1+r}{3r}} \mathbb{P}(|X|^r \tilde{q})^{\frac{1}{r\tilde{q}}} \mathbb{P}(|\tilde{\kappa}_n(X) - \kappa(X)|^{r\tilde{q}})^{\frac{1}{r\tilde{q}}} + 3\mathbb{P}(|\tilde{\kappa}_n(X) - \kappa(X)|^r)^{\frac{1}{r}}.
\end{aligned}$$

Considering that, from (A0), $\mathbb{P}(|X|^r D(X)^r \mathbf{1}_{|\beta_0^T X| \leq |\hat{\beta}_n - \beta_0||X|}) \leq |\beta_0 - \hat{\beta}_n|$ and $\mathbb{P}(|\tilde{\kappa}_n(X) - \kappa(X)|^{r\tilde{q}})^{\frac{1}{r\tilde{q}}} = o_{\mathbf{P}}\left(n^{-\frac{2+r}{3r}}\right)$ we can conclude from that (ii) holds. A similar argument gives (iii). \square

B.2 Some properties of the multivariate empirical distribution

Here we discuss some properties of the empirical distribution function that will allow us to apply the results of Section 3 to prove the consistency of some of our bootstrap procedures. Consider a sequence $(X_n, U_n)_{n=1}^{\infty} \stackrel{i.i.d.}{\sim} \mathbb{P}$ on the probability space $(\Omega, \mathcal{A}, \mathbf{P})$ and let μ_n be the EDF of X_1, \dots, X_n . Also, consider the classes of functions $\mathcal{F}_{n,R}$ as defined in Section 3, with their envelopes $F_{n,R}$, replacing β_n with $\hat{\beta}_n$. Then, we have the following result.

Lemma B.3 *Let $(\hat{\kappa}_n)_{n=1}^{\infty}$ be a sequence satisfying (i) in the description of the fixed design scheme (see Section 2.2.2) and consider a decreasing sequence $(\epsilon)_{n=1}^{\infty}$ with $\epsilon_n \downarrow 0$ and $\epsilon_n n^{\frac{1}{3r}} \rightarrow \infty$. Then, there are constants $C, R_0 > 0$ such that the following hold:*

- (a) *the sequence $(\mu_n)_{n=1}^{\infty}$ is tight with probability one;*
- (b) *for every sequence $(\mu_{n_k})_{k=1}^{\infty}$ there is a subsubsequence $(\mu_{n_{k_s}})_{s=1}^{\infty}$ such that*

$$\sup_{0 \leq R \leq R_0} \{ |(\mu_{n_{k_s}} - \mathbb{P})(F_{n_{k_s}, R}^2) | \} \xrightarrow{a.s.} 0;$$

(c) for every sequence $(\mu_{n_k})_{k=1}^\infty$ there is a subsubsequence $(\mu_{n_{k_s}})_{s=1}^\infty$ such that, with probability one, the following inequality holds for all $0 < R \leq R_0$,

$$\sup_{\substack{|\alpha - \hat{\beta}_{n_{k_s}}| \vee |\beta - \hat{\beta}_{n_{k_s}}| \leq R \\ |\alpha - \beta| \leq R}} |(\mu_{n_{k_s}} - \mathbb{P})(\hat{\kappa}_{n_{k_s}}(X)(\mathbf{1}_{\alpha^T X \geq 0} - \mathbf{1}_{\beta^T X \geq 0}))| \leq CRn_{k_s}^{-1/3} \epsilon_{n_{k_s}};$$

(d) for every sequence $(\mu_{n_k})_{k=1}^\infty$ there is a subsubsequence $(\mu_{n_{k_s}})_{s=1}^\infty$ such that, with probability one, we have that the following inequality holds for all $0 < R \leq R_0$,

$$\sup_{\substack{|\alpha - \hat{\beta}_{n_{k_s}}| \vee |\beta - \hat{\beta}_{n_{k_s}}| \leq R \\ |\alpha - \beta| \leq R}} |(\mu_{n_{k_s}} - \mathbb{P})(\mathbf{1}_{\alpha^T X \geq 0} - \mathbf{1}_{\beta^T X \geq 0})| \leq CRn_{k_s}^{-1/3} \epsilon_{n_{k_s}}.$$

Proof: The collection of cells of the form $(a, b]$ in \mathbb{R}^d with $a, b \in \overline{\mathbb{R}}^d$ is VC (see example 2.6.1, page 135 in [Van der Vaart and Wellner \(1996\)](#)). It follows that that class is (strongly) Glivenko-Cantelli and thus, μ_n converges weakly to μ with probability one and hence (a) is true. On the other hand, since there are $\beta_R, \alpha_R \in \mathbb{R}^d$ such that

$$\begin{aligned} F_{n,R} &= \mathbf{1}_{\beta_R^T x \geq 0 > \alpha_R^T x} + \mathbf{1}_{\alpha_R^T x \geq 0 > \beta_R^T x}, \\ &= (\mathbf{1}_{\beta_R^T x \geq 0} \wedge \mathbf{1}_{0 > \alpha_R^T x}) \vee (\mathbf{1}_{\alpha_R^T x \geq 0} \wedge \mathbf{1}_{0 > \beta_R^T x}), \end{aligned} \quad (4)$$

a similar argument to that used to show that \mathcal{F}_1 and \mathcal{F}_2 are VC in the proof of Lemma 3.2 applies here prove that $\{F_{n,R} : 0 \leq R \leq R_0\}$ is VC. Thus, (b) follows from a suitable application of the maximal inequality 3.1 of [Kim and Pollard \(1990\)](#). Now, fix $R > 0$ small enough. The classes $\mathcal{F}_{n,R}$ and $\{\kappa\psi\}_{\psi \in \mathcal{F}_{n,R}}$ are both VC with envelope $F_{n,R}$. Considering the fact that $\hat{\beta}_n - \beta_0 = O_{\mathbf{P}}(1)$ with probability tending to one, we have that $|\alpha - \beta_0| \vee |\beta - \beta_0| \leq 2R$ for all $\alpha, \beta \in \mathcal{S}^{d-1}$ with $|\alpha - \beta| \leq R$ and $|\alpha - \hat{\beta}_n| \vee |\beta - \hat{\beta}_n| \leq R$. Thus, a suitable application of inequality 3.1 in [Kim and Pollard \(1990\)](#) shows that with probability tending to one,

$$\sup_{\substack{|\alpha - \hat{\beta}_n| \vee |\beta - \hat{\beta}_n| \leq R \\ |\alpha - \beta| \leq R}} |(\mu_n - \mathbb{P})(\mathbf{1}_{\alpha^T X \geq 0} - \mathbf{1}_{\beta^T X \geq 0})| \leq \sup_{\substack{|\alpha - \beta_0| \vee |\beta - \beta_0| \leq R \\ |\alpha - \beta| \leq R}} |(\mu_n - \mathbb{P})(\mathbf{1}_{\alpha^T X \geq 0} - \mathbf{1}_{\beta^T X \geq 0})| = O_{\mathbf{P}}(n^{-1/2}).$$

Hence, for every R small enough there is a subsequence for which (d) holds. To obtain a subsequence for which (d) holds almost surely for all R smaller than some R_0 one needs to apply the previous argument to all rational R 's smaller than R_0 and then use Cantor's diagonal method.

It remains to prove (c). An argument like the one used for (d) proves that (c) is true with κ replacing $\hat{\kappa}_n$. Finally, the equation

$$(\mu_n - \mathbb{P})(\hat{\kappa}_n(X)(\mathbf{1}_{\alpha^T X \geq 0} - \mathbf{1}_{\beta^T X \geq 0})) = \mu_n((\hat{\kappa}_n(X) - \kappa(X))(\mathbf{1}_{\alpha^T X \geq 0} - \mathbf{1}_{\beta^T X \geq 0})) + \\ (\mu_n - \mathbb{P})(\kappa(X)(\mathbf{1}_{\alpha^T X \geq 0} - \mathbf{1}_{\beta^T X \geq 0})) - \mathbb{P}((\hat{\kappa}_n(X) - \kappa(X))(\mathbf{1}_{\alpha^T X \geq 0} - \mathbf{1}_{\beta^T X \geq 0})),$$

together with Hölder's inequality and the fact that $\mu_n(|\hat{\kappa}_n - \kappa|^r)^{1/r} \vee \mathbb{P}(|\hat{\kappa}_n - \kappa|^r)^{1/r} = o_{\mathbf{P}}\left(n^{-\frac{r+1}{3r}}\right)$ give the result. \square

References

- Abrevaya, J. and Huang, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica*, 73:1175–1204.
- Aubin, J. and Frankowska, H. (2009). *Set-Valued Analysis*. Birkhäuser, Boston, MA, USA.
- Bickel, P. and Freedman, D. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.*, 9:1196–1217.
- Cavanagh, C. L. (1987). Limiting behavior of estimators defined by optimization. unpublished manuscript.
- Cosslett, S. R. (1983). Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica*, 51:765–782.
- Delgado, M., Rodríguez-Poo, J., and Wolf, M. (2001). Subsampling inference in cube root asymptotics with an application to manski's maximum score estimator. *Econom. Lett.*, 73:241–250.
- Han, A. K. (1987). Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *J. Econometrics*, 35:303–316.
- Horowitz, J. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica*, 60:505–531.

- Horowitz, J. (2002). Bootstrap critical values for tests based on the smoothed maximum score estimator. *J. Econometrics*, 111:141–167.
- Horowitz, J. and Härdle, W. (1996). Direct semiparametric estimation of a single-index model with discrete covariates. *Jour. Amer. Statist. Assoc.*, 91:1632–1640.
- Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.*, 18:191–219.
- Lee, S. and Pun, M. (2006). On m out of n bootstrapping for nonstandard estimation with nuisance parameters. *J. Amer. Statist. Assoc.*, 101:1185–1197.
- Manski, C. (1975). Maximum score estimation of the stochastic utility model of choice. *J. Econometrics*, 3:205–228.
- Manski, C. (1985). Semiparametric analysis of a discrete response. *J. Econometrics*, 27:313–333.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, ed. by P. Zarembka. New York: Academic Press:105–142.
- Meyer, C. D. (2001). *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, PA, USA.
- Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics, United States of America.
- Powell, J. L., S. J. H. and Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, 57:1403–1430.
- Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, New York, NY.
- Sen, B., Banerjee, M., and Woodroffe, M. (2010). Inconsistency of bootstrap: The grenander estimator. *Ann. Statist.*, 38:1953–1977.
- Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer-Verlag, New York, USA.

- Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica*, 61:123–138.
- Singh, K. (1981). On asymptotic accuracy of efron's bootstrap. *Ann. Statist.*, 9:1187–1195.
- Stone, C. (1982). Optimal global rates of convergence for nonparametric regression estimators. *Ann. Statist.*, 10:1040–1053.
- Van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York, NY, USA.