



General M -Estimator Processes and their m out of n Bootstrap with Functional Nuisance Parameters

Salim Bouzebda¹ · Issam Elhattab^{1,2} · Anouar Abdeldjaoued Ferfache¹

Received: 21 March 2021 / Revised: 4 January 2022 / Accepted: 25 May 2022 /
Published online: 16 June 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

In the present paper, we consider the problem of the estimation of a parameter θ , in Banach spaces, maximizing some criterion function which depends on an unknown nuisance parameter h , possibly infinite-dimensional. The classical estimation methods are mainly based on maximizing the corresponding empirical criterion by substituting the nuisance parameter by a nonparametric estimator. We show that the M -estimators converge weakly to maximizers of Gaussian processes under rather general conditions. The conventional bootstrap method fails in general to consistently estimate the limit law. We show that the m out of n bootstrap, in this extended setting, is weakly consistent under conditions similar to those required for weak convergence of the M -estimators. The aim of this paper is therefore to extend the existing theory on the bootstrap of the M -estimators. Examples of applications from the literature are given to illustrate the generality and the usefulness of our results. Finally, we investigate the performance of the methodology for small samples through a short simulation study.

Keywords Gaussian process · M -estimation · Empirical process · m out of n of bootstrap · Asymptotic distribution · Nuisance parameter · Semiparametric estimation · Non standard distribution · Missing data

AMS Subject Classification Primary: 62G05 · 60F17 · Secondary: 60F05 · 62G09 · 62G20 · 62H10 · 60F15

✉ Salim Bouzebda
salim.bouzebda@utc.fr

Issam Elhattab
issam.elhattab@gmail.com

Anouar Abdeldjaoued Ferfache
ferfache.abdeldjaoued@gmail.com

¹ LMAC (Laboratory of Applied Mathematics of Compiègne), Université de technologie de Compiègne, Compiègne, France

² École Nationale de Commerce et de Gestion (ENCG), Hassan II University of Casablanca, Casablanca, Morocco

1 Introduction

The semiparametric modeling has proved to be a flexible tool and provided a powerful statistical modeling framework in a variety of applied and theoretical contexts [refer to Pfanzagl (1990), Bickel et al. (1993), van der Vaart and Wellner (1996), van de Geer (2000), and Kosorok (2008)]. An important work to be cited is the paper of Pakes and Pollard (1989), where a general central limit theorem is proved for estimators defined by minimization of the length of a vector-valued, random criterion function with no smoothness assumptions. The last reference was extended in different settings, among many others, by Pakes and Olley (1995), Chen et al. (2003), Zhan (2002). Recall that the semiparametric models are statistical models where at least one parameter of interest is not Euclidean. The term “*M*-estimation” refers to a general method of estimation, where the estimators are obtained by maximizing (or minimizing) certain criterion functions. The most widely used *M*-estimators include maximum likelihood (MLE), ordinary least-squares (OLS), and least absolute deviation estimators. Notice that the major practical problem of maximum likelihood estimators is the lack of robustness, while many robust estimators achieve robustness at some cost in first-order efficiency. The appeal of the *M*-estimation method is that in addition to the statistical efficiency of the estimators when the parametric model is correctly specified, these estimators are also robust to contamination when the objective function is appropriately chosen. Throughout the available literature, investigations on the asymptotic properties of the *M*-estimators, as well as the relevant test statistics, have privileged the parametric case. However, in practice, we need more flexible models that contain both parametric and nonparametric components. This paper concentrates on this specific problem. To formulate the problem that we will treat in this paper, we need the following notation. Let $\mathcal{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ be n independent copies of a random element \mathbf{X} in a probability space $(\mathcal{S}, \mathcal{A}, \mathbb{P})$. For a Banach spaces \mathcal{B} and \mathcal{H} equipped with a norm $\|\cdot\|$ and a metric denoted by $d_{\mathcal{H}}(\cdot, \cdot)$ respectively, let $\mathcal{M}_{\Theta, \mathcal{H}}$ be a class of Borel measurable functions $\mathbf{m}_{\theta, h} : \mathcal{S} \rightarrow \mathbb{R}$, indexed by θ over some parameter space $\Theta \subset \mathcal{B}$ and $h \in \mathcal{H}$, where θ is the parameter of interest and h^0 the true value of h consists of nuisance parameter. We define the empirical measure to be

$$\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{\mathbf{X}_i},$$

where, for $\mathbf{x} \in \mathcal{S}$, $\delta_{\mathbf{x}}$ is the measure that assigns mass 1 at \mathbf{x} and zero elsewhere. Let $f(\cdot)$ be a real valued measurable function, $f : \mathcal{S} \rightarrow \mathbb{R}$. In the modern theory of the empirical it is usual to identify \mathbb{P} and \mathbb{P}_n with the mappings given by

$$f \rightarrow \mathbb{P}f = \int f d\mathbb{P}, \quad \text{and} \quad f \rightarrow \mathbb{P}_n f = \int f d\mathbb{P}_n = \frac{1}{n} \sum_{k=1}^n f(\mathbf{X}_i).$$

The *M*-estimand of interest θ_0 and its corresponding *M*-estimator θ_n are assumed to be well-separated maximizers of the processes $\left\{ \mathbb{P} \mathbf{m}_{\theta, h^0} : \theta \in \Theta \right\}$ and $\left\{ \mathbb{P}_n \mathbf{m}_{\theta, \hat{h}} : \theta \in \Theta \right\}$ for a given consistent sequence of estimators \hat{h} for h^0 , respectively. Under suitable entropy conditions on $\mathcal{M}_{\Theta, \mathcal{H}}$ (defined below) and moment conditions on its envelope, we show that there exist norming sequences $\{\alpha_n\}$ and $\{r_n\}$ such that the random process $\left\{ \alpha_n \mathbb{P}_n (\mathbf{m}_{\theta_0 + \gamma / r_n, \hat{h}} - \mathbf{m}_{\theta_0, \hat{h}}) : \gamma \in K \right\}$ converges weakly, in the sense of Hoffmann-Jørgensen (1991), see van der Vaart and Wellner (1996), in particular their Definition

1.3.3., to the process $\{\mathbb{Z}(\gamma) : \gamma \in K\}$, for each closed bounded subset $K \subset \mathcal{B}$. It follows by an argmax continuous mapping theorem, refer to Kosorok (2008) in particular Chapter 14, that $r_n(\hat{\theta}_n - \theta_0)$ converges weakly to $\arg \max_{\gamma} \mathbb{Z}(\gamma)$. The latter weak limit has a complicated form in general and does not permit explicit computation. It would therefore be of interest to estimate the sampling distribution of $r_n(\hat{\theta}_n - \theta_0)$ by the bootstrap for inferencing purposes. Bootstrap samples were introduced and first investigated in Efron (1979). Since this seminal paper, bootstrap methods have been proposed, discussed, investigated and applied in a huge number of papers in the literature. Being one of the most important ideas in the practice of statistics, the bootstrap also introduced a wealth of innovative probability problems, which in turn formed the basis for the creation of new mathematical theories. The bootstrap can be described briefly as follows. Let $T(\mathbb{F})$ be a functional of an unknown distribution function $\mathbb{F}(\cdot)$, $\mathbf{X}_1, \dots, \mathbf{X}_n$ a sample from $\mathbb{F}(\cdot)$, and $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ an independent and identically distributed [i.i.d.] sample with common distribution given by the empirical distribution $\mathbb{F}_n(\cdot)$ of the original sample. The distribution of $\{T(\mathbb{F}_n) - T(\mathbb{F})\}$ is then approximated by that of $\{T(\mathbb{F}_n^*) - T(\mathbb{F}_n)\}$ conditionally on $\mathbf{X}_1, \dots, \mathbf{X}_n$, with $\mathbb{F}_n^*(\cdot)$ being the empirical distribution of $\mathbf{Y}_1, \dots, \mathbf{Y}_n$. The key idea behind the bootstrap is that if a sample is representative of the underlying population, then one can make inferences about the population characteristics by resampling from the current sample. The asymptotic theory of the bootstrap with statistical applications has been reviewed in the books among others Efron and Tibshirani (1993) and Shao and Tu (1995). Chernick (2008), Davison and Hinkley (1997), van der Vaart and Wellner (1996), Hall (1992) and Kosorok (2008). A major application for an estimator is in the calculation of confidence intervals. By far the most favored confidence interval is the standard confidence interval based on a normal or a Student t -distribution. Such standard intervals are useful tools, but based on an approximation that can be quite inaccurate in practice. Bootstrap procedures are an attractive alternative. One way to look at them is as procedures for handling data when one is not willing to make assumptions about the parameters of the populations from which one sampled. The most that one is willing to assume that the data are a reasonable representation of the population from which they come. One then resamples from the data and draws inferences about the corresponding population and its parameters. The resulting confidence intervals have received the most theoretical study of any topic in the bootstrap analysis. Roughly speaking, it is known that the bootstrap works in the i.i.d. case if and only if the central limit theorem holds for the random variable under consideration. For further discussion we refer the reader to the landmark paper by Giné and Zinn (1989). It is worth noticing that some special examples reveal that the conventional bootstrap based on resamples of size n breaks down; see, for example, Bose and Chatterjee (2001) and El Bantli (2004). We focus on a modified form of bootstrap methods, known as the m out of n bootstrap, with a view to remedy the problem of inconsistency. The m out of n scheme modifies the conventional scheme by drawing bootstrap resamples of size $m = o(n)$. See, for example, Bickel et al. (1997) for a review of this technique in a variety of contexts. For more recent references on the bootstrap one can refer to Bouzebda (2010), Bouzebda and Limnios (2013), Bouzebda et al. (2018), Alvarez-Andrade and Bouzebda (2013, 2015, 2019) and the reference therein. Denote by $\hat{\theta}_m$ the M -estimator calculated from a bootstrap resample of size m . Weak convergence in probability of the conditional distribution of $r_m(\hat{\theta}_m - \hat{\theta}_n)$ to the distribution of $\arg \max_{\gamma} \mathbb{Z}(\gamma)$ is established under essentially similar conditions for weak convergence of $r_n(\hat{\theta}_n - \theta_0)$, provided that $m = o(n)$, $m \rightarrow \infty$ and $a_m^2 m^{-1/2} \log n / \log(n/m + 1) = o(1)$ for a fixed sequence $\{a_m\}$ depending on the size of the envelope for $\mathcal{M}_{\theta, \mathcal{H}}$. The asymptotic properties of $\hat{\theta}_n$ have been established by, among many others, Bose and Chatterjee (2001) and El Bantli (2004), under appropriate concavity or differentiability conditions. Empirical

process methods are instrumental tools for evaluating the large sample properties of estimators based on semiparametric models, including consistency, distributional convergence, and validity of the bootstrap. In particular, modern empirical process theory provides a more general approach to theoretical investigation of general M -estimators; see, for example, Dudley (1999), Kim and Pollard (1990), Pollard (1985), van de Geer (2000) and van der Vaart and Wellner (1996). Most results obtained thus far are, however, restricted to the cases where the Gaussian process Z has either quadratic mean function or quadratic covariance function. In order to establish stronger results which cover cases where the covariance and mean functions of Z may take on more general structures, we will use the empirical process approach. Applications of the bootstrap to M -estimation have been investigated deeply in the literature extensively. Relevant theoretical results concern mostly M -estimators with $r_n = n^{1/2}$ and asymptotically Gaussian limits. The most common technique for studying bootstrap M -estimators is the linearization which can not be used in a nonstandard setting. Under standard regularity conditions, the Edgeworth expansions for bootstrap distributions of finite-dimensional M -estimators are Lahiri (1992). Under a weak form of differentiability condition, Arcones and Giné (1992) investigated bootstrapping finite-dimensional $n^{1/2}$ -consistent M -estimators and established an almost sure bootstrap central limit theorem. An in-probability bootstrap central limit theorem for possibly infinite-dimensional Z -estimators is investigated by Wellner and Zhan (1996). In the setting of the nonregular vector-valued M -estimators obtained from \mathbf{m}_θ concave in θ , Bose and Chatterjee (2001) investigated a weighted form of the bootstrap including the m out of n bootstrap is a special case. The M -estimation for linear models under nonstandard conditions was considered by El Bantli (2004), and proved that the m out of n bootstrap is consistent but the conventional bootstrap is not in general. The Bose and Chatterjee (2001) and El Bantli (2004) results are restricted to the case where Z has a quadratic covariance function, concavity and differentiability assumptions. Lee and Pun (2006) prove m out of n bootstrap consistency for vector-valued M -estimators under twice-differentiability of the process $\mathbb{P}\mathbf{m}_\theta$, where θ may contain a subvector of nuisance parameters, in which case the process Z has a quadratic mean function. Lee (2012) gives the general result of m out of n bootstrap of M -estimators without the presence of a nuisance parameter. Under nonstandard conditions, Lee and Yang (2020) proposed a one-dimensional pivot derived from the criterion function, and prove that its distribution can be consistently estimated by the m out of n bootstrap, or by a modified version of the perturbation bootstrap. They provide a new method for constructing confidence regions which are transformation equivariant and have shapes driven solely by the criterion function.

The main purpose of the present work is to consider a general framework of non-smooth semi-parametric M -estimators extending the setting of Lee (2012) to the \mathcal{B} -valued M -estimators in presence of nuisance parameter where the rate of convergence of the nuisance parameter may be different of that of the parameter of interest. More precisely, we consider the m out of n bootstrapped version of the M -estimator investigated in Delsol and Van Keilegom (2020), where these authors showed that, their M -estimator converges weakly to some process which is composed on Gaussian process and some deterministic continuous function, which is harder to evaluate for practical use. For that we propose in this paper as a solution to this problem the m out of n bootstrap. We mention at this stage that parameter θ , in the present paper, belongs to some Banach space which is different from the last mentioned work where the parameter of interest is euclidean. Hence, we restate the results of Delsol and Van Keilegom (2020) under more general conditions. The main aim of the present paper is to provide a first full theoretical justification of the m out of n bootstrap consistency of M -estimators with non-smooth criterion

functions of the parameters and gives the consistency rate together with the asymptotic distribution of the parameters of interest θ_0 . This requires the effective application of large sample theory techniques, which were developed for the empirical processes. The Lee (2012) results are not directly applicable here since the estimation procedures depend on some nuisance parameters. These results are not only useful in their own right but essential for the derivation of our asymptotic results.

The paper is organized as follows. Section 2 introduces the notation and assumptions. Section 3 states the main theorems. Though our main objective in the paper is theoretical, we provide in Sect. 4 Monte Carlo simulations of simulations to look at the method’s performance in practice. Some concluding remarks are given in Sect. 5. All proofs are gathered in Sect. 6. In the Appendix we apply our theorems and prove as corollaries new m out of n bootstrap consistency results for three examples.

2 Notation

We abuse notation slightly by identifying the underlying probability space $(S, \mathcal{A}, \mathbb{P})$ with the product space $(S^\infty, \mathcal{A}^\infty, \mathbb{P}^\infty) \times (Z, \mathcal{C}, P)$. Now $\mathbf{X}_1, \dots, \mathbf{X}_n$ are equal to the coordinate projections on the first n coordinates. All auxiliary variables, assumed to be independent of the \mathbf{X}_i , depend only on the last coordinate. We will use the usual notation of the empirical processes of van der Vaart and Wellner (1996). Let \mathbb{Q} denote some signed measure on S . Let \mathcal{F} be a class of measurable functions $f : S \rightarrow \mathbb{R}$. Define

$$\|\mathbb{Q}f\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\mathbb{Q}f|.$$

For any $r \geq 1$, denote by $L^r(\mathbb{Q})$ the class of measurable functions $f : S \rightarrow \mathbb{R}$ with

$$\int |f|^r d\mathbb{Q} < \infty,$$

where \mathbb{Q} is a probability measure. The $L^r(\mathbb{Q})$ -norm $\|\cdot\|_{\mathbb{Q},r}$ is defined by

$$\|f\|_{\mathbb{Q},r} = \left(\int |f|^r d\mathbb{Q} \right)^{1/r},$$

for $f \in L^r(\mathbb{Q})$. The essential supremum of $f \in L^\infty(\mathbb{Q})$ is denoted by $\|f\|_{\mathbb{Q},\infty}$.

The covering number $N(\epsilon, \mathcal{F}, L^r(\mathbb{Q}))$ of a function class $\mathcal{F} \subset L^r(\mathbb{Q})$ is computed with respect to the $L^r(\mathbb{Q})$ -norm for radius $\epsilon > 0$. To be more precise, $N(\epsilon, \mathcal{F}, L^r(\mathbb{Q}))$ is the minimum number of balls $\{g : \|g - h\|_{\mathbb{Q},r} < \epsilon\}$ of radius ϵ covering \mathcal{F} .

For some random element \mathbf{Z} , the probability measure induced by \mathbf{Z} is denoted by $\mathbb{P}_{\mathbf{Z}}$, conditional on all other variables. The empirical process is defined to be

$$\mathbb{G}_n = n^{1/2}(\mathbb{P}_n - \mathbb{P}).$$

The outer and inner probability measures derived from \mathbb{P} are designated by \mathbb{P}^* and \mathbb{P}_* , respectively. Outer and inner probability measures to be understood in the sense used in the monograph by van der Vaart and Wellner (1996), in particular their definitions in page 6. Let T be any map from the underlying probability space to the extended real line $\overline{\mathbb{R}}$. The minimal measurable majorant and maximal measurable minorant of T are denoted by T^* and T_* , respectively. For any subset B of the probability space, by similar notation, its indicator function

satisfies $\mathbb{1}_{B^*} = \mathbb{1}_B^*$ and $\mathbb{1}_{B_*} = (\mathbb{1}_B)_*$. We draw randomly with replacement from \mathcal{X} independent bootstrap observations $\mathbf{Y}_1, \dots, \mathbf{Y}_m$. Let us define

$$\widehat{\mathbb{P}}_m = m^{-1} \sum_{i=1}^m \delta_{\mathbf{Y}_i},$$

so that

$$\widehat{\mathbb{P}}_m = \sum_{i=1}^m W_i \delta_{\mathbf{X}_i},$$

where $mW = m(W_1, \dots, W_m)$ is a multinomial vector with m trials and parameters (n^{-1}, \dots, n^{-1}) , independent of the \mathbf{X}_i . The probability measure induced by bootstrap resampling conditional on \mathcal{X} is denoted by P_W . Let us define the bootstrapped empirical process by

$$\widehat{\mathbb{G}}_m = m^{1/2} \left(\widehat{\mathbb{P}}_m - \mathbb{P}_n \right).$$

Let T_n denote a sequence of maps. Let \mathbb{D} be a metric space. Let T be a \mathbb{D} -valued measurable map from the underlying probability. If T_n is bounded in outer probability, we will write $T_n = O_{\mathbb{P}^*}(1)$, in a similar way, if T_n converges in outer probability to zero, we will write $T_n = o_{\mathbb{P}^*}(1)$. Assume that

$$\lim_{M \rightarrow \infty} \liminf_{n \rightarrow \infty} \mathbb{P}_W \left\{ \|T_n\| < M \right\}_* = 1. \tag{1}$$

If (1) holds along almost every sequence $\mathbf{X}_1, \mathbf{X}_2, \dots$, we write $T_n = O_{\mathbb{P}_W^*}(1)$ a.s. (almost surely). If for any subsequence $\{T_{n'}\}$, there exists a further subsequence $\{T_{n''}\}$ with $T_{n''} = O_{\mathbb{P}_W^*}(1)$ a.s., we write $T_n = O_{\mathbb{P}_W^*}(1)$ i.p. (in probability). We write $T_n = o_{\mathbb{P}_W^*}(1)$ a.s., if, for any $\epsilon > 0$, we have

$$\mathbb{P}_W \left\{ \|T_n\| > \epsilon \right\}^* \rightarrow 0, \text{ as } n \rightarrow \infty \tag{2}$$

almost surely. We write $T_n = o_{\mathbb{P}_W^*}(1)$ i.p., in the case when the convergence (2) is in probability. The weak convergence of T_n to T , in the sense of Hoffmann-Jørgensen (1991), is denoted by $T_n \Rightarrow T$. The space of \mathbb{D} -valued functions in \mathbb{R} bounded by 1 in the Lipschitz norm is denoted by $BL_1(\mathbb{D})$. The conditional weak convergence of T_n to a separable T in \mathbb{D} is characterized by the condition

$$\sup_{f \in BL_1(\mathbb{D})} \left| \mathbb{P}_W^* f(T_n) - \mathbb{P} f(T) \right| \rightarrow 0. \tag{3}$$

In the case of the convergence (3) is in outer probability, we will write $T_n \Rightarrow T$ i.p., in a similar way, if it is outer almost sure, we write $T_n \Rightarrow T$ a.s.

Define $\mathcal{M}_{S, \mathcal{H}} = \{ \mathbf{m}_{\theta, h} : \theta \in S, h \in \mathcal{H} \} \subset \mathcal{M}_{\Theta, \mathcal{H}}$, where $S \subset \Theta$. For any $\delta, \delta_1, \eta > 0$, let us denote by $\mathcal{M}_{\delta, \delta_1}(\eta)$ and $\mathcal{M}_{\delta, \delta_1}$ the class of functions

$$\mathcal{M}_{\delta, \delta_1} = \left\{ \mathbf{m}_{\theta, h} - \mathbf{m}_{\theta_0, h} : \|\theta - \theta_0\| \leq \delta, d_{\mathcal{H}}(h, h^0) \leq \delta_1, \theta \in \Theta, h \in \mathcal{H} \right\},$$

$$\mathcal{M}_{\delta, \delta_1}(\eta) = \left\{ \mathbf{m}_{\theta, h} - \mathbf{m}_{\psi, h} : \|\theta - \psi\| < \eta, \|\theta - \theta_0\| \vee \|\psi - \theta_0\| < \delta, d_{\mathcal{H}}(h, h^0) \leq \delta_1, \theta, \psi \in \Theta, h \in \mathcal{H} \right\}.$$

The envelope function of $\mathcal{M}_{\delta, \delta_1}$ is denoted by M_{δ, δ_1} . For each $\psi \in \mathcal{B}$ and $h \in \mathcal{H}$ with $\theta_0 + \psi \in \Theta$, define $\tilde{\mathbf{m}}_{\psi, h} = \mathbf{m}_{\theta_0 + \psi, h} - \mathbf{m}_{\theta_0, h}$. For any $\mathcal{T} \subset \mathcal{B}$, the class of bounded functions from \mathcal{T} to \mathbb{R} is denoted by $l^\infty(\mathcal{T})$, equipped with the sup norm. In the sequel, for all $x \in \mathcal{S}$ and closed bounded $K \subset \Theta$, assume that

$$\sup_{\theta \in K, h \in \mathcal{H}} |\mathbf{m}_{\theta, h}(x) - \mathbb{P}\mathbf{m}_{\theta, h}| < \infty.$$

In the sequel, we denote by C a positive constant that may be different from line to line. The choice of the bootstrap sample size m is theoretically governed by **(A1)** and **(C4)**. The above conditions are typically satisfied by taking $m \propto n^c$, for some sufficiently small $c \in (0, 1)$. Empirical determination of m has long been an important problem which has not yet been fully resolved, for more comments see Remark 3.12 below.

3 Main Results

In this section, we present four main theorems, each of independent interest, which lead eventually to weak convergence of $r_n(\theta_n - \theta_0)$ and in-outer-probability m out of n bootstrap consistencies in the context of general M -estimation by applying the argmax theorem in van der Vaart and Wellner (1996) and in Lee (2012) respectively. Let us recall the basic idea. If the argmax functional is continuous with respect to some metric on the space of the criterion functions, then convergence in distribution of the criterion functions will imply the convergence in distribution of their points of maximum, the M -estimators, to the maximum of the limit criterion function. First, we establish consistency of θ_n and θ_m for θ_0 by the following theorem.

3.1 Consistency

In our analysis, we consider the following assumptions. Assume that the sequence of positive constants $r_n \uparrow \infty$, for some fixed $\nu > 1$ and for some function $\ell : (0, \infty) \rightarrow [0, \infty)$ which is slowly varying at ∞ .

(A1) $\mathbb{P}(\hat{h} \in \mathcal{H}) \rightarrow 1$ as $n \rightarrow \infty$ and $d_{\mathcal{H}}(\hat{h}, h^0) \xrightarrow{\mathbb{P}^*} 0$.

(A2) $\mathcal{M}_{\Theta, \mathcal{H}}$ is Glivenko-Cantelli:

$$\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{M}_{\Theta, \mathcal{H}}} = o_{\mathbb{P}^*}(1).$$

(A3) $\lim_{d_{\mathcal{H}}(h, h_0) \rightarrow 0} \sup_{\theta \in \Theta} |\mathbb{P}\mathbf{m}_{\theta, h} - \mathbb{P}\mathbf{m}_{\theta, h^0}| = 0$.

(A4) The parameter of interest θ_0 lies in the interior of Θ and satisfies, for every open \mathcal{O} containing θ_0 ,

$$\mathbb{P}\mathbf{m}_{\theta_0, h^0} > \sup_{\theta \in \mathcal{O}} \mathbb{P}\mathbf{m}_{\theta, h^0}.$$

(A5) The M -estimator θ_n satisfies $\mathbb{P}_n \mathbf{m}_{\theta_n, \hat{h}} \geq \mathbb{P}_n \mathbf{m}_{\theta_0, \hat{h}} - R_n$, with

$$r_n^\nu \ell(r_n) R_n = o_{\mathbb{P}^*}(1).$$

(AB1) $m = m_n \rightarrow \infty$, $m = o(n)$ and $r_m^\nu \ell(r_m) = o(r_n^\nu \ell(r_n))$.

(AB2) $d_{\mathcal{H}}(\hat{h}_m, h^0) = o_{\mathbb{P}_w^*}(1)$ i.p.

(AB3) The m out of n bootstrap M -estimator θ_m satisfies $\hat{\mathbb{P}}_m \mathbf{m}_{\theta_m, \hat{h}_m} \geq \hat{\mathbb{P}}_m \mathbf{m}_{\theta_0, \hat{h}_m} - \hat{R}_n$, with

$$r_m^\nu \ell(r_m) \hat{R}_n = o_{\mathbb{P}_w^*}(1), \text{ i.p.}$$

Remark 3.1

- (i) Assumption (A2) fulfilled under some entropy and moment conditions; see for example, Theorem 2.4.3, (p.123) of van der Vaart and Wellner (1996).
- (ii) Assumption (A3) is automatically hold for example if; there exist function $\mathfrak{G}(\cdot)$ such that for any h in the neighborhood of h^0 and any $\theta \in \Theta$, we have:

$$|\mathbf{m}(X_i, \theta, h) - \mathbf{m}(X_i, \theta, h^0)| \leq \mathfrak{G}(X_i) d_{\mathcal{H}}(h, h^0).$$

The function $\mathfrak{G}(\cdot)$ satisfies;

$$\mathbb{P} \mathfrak{G}(X) < \infty,$$

or the function $h \mapsto \mathbf{m}(x, \theta, h)$ is Lipschitz uniformly over x and θ .

- (iii) Assumptions (A5) and (AB3) are trivially fulfilled when

$$\mathbb{P}_n \mathbf{m}_{\theta_n, \hat{h}} \geq \sup_{\theta \in \Theta} \mathbb{P}_n \mathbf{m}_{\theta, \hat{h}} - R_n,$$

and

$$\hat{\mathbb{P}}_m \mathbf{m}_{\theta_m, \hat{h}_m} \geq \sup_{\theta \in \Theta} \hat{\mathbb{P}}_m \mathbf{m}_{\theta, \hat{h}_m} - \hat{R}_n,$$

respectively, which allows to deal with approximations of the value that actually maximizes $\theta \mapsto \mathbb{P}_n \mathbf{m}_{\theta, \hat{h}}$ and maximizes $\theta \mapsto \hat{\mathbb{P}}_m \mathbf{m}_{\theta, \hat{h}_m}$ respectively.

- (iv) Assumption (AB2) poses no difficulty in practice and is met trivially by, for example, setting $\hat{h}_m = \hat{h}$.
- (v) For the finite-dimensional θ , (A5) and (AB3) can be achieved by a global maximization of the processes $\mathbb{P}_n \mathbf{m}_{\theta, \hat{h}}$ and $\hat{\mathbb{P}}_m \mathbf{m}_{\theta, \hat{h}_m}$, in this situation $R_n = \hat{R}_n = 0$. For the infinite-dimensional θ , the maximization of the processes may be very complex or not practically feasible. To circumvent this, we need sophisticated algorithms to construct θ_n and θ_m fulfilling (A5) and (AB3).
- (vi) Finally, it's possible to replace the following assumptions (A2) and (A4) by:

(A1') For every compact $K \subset \Theta$, $\mathcal{M}_{K, \mathcal{H}}$ is Glivenko-Cantelli.

(A2') The map $\theta \mapsto \mathbb{P} \mathbf{m}_{\theta, h^0}$ is upper semicontinuous with a unique maximum at θ_0 .

(A3') θ_n is uniformly tight.

(AB1') θ_m is uniformly tight i.p.

Theorem 3.2

(i) Assume **(A1)**–**(A5)**. Then

$$\theta_n - \theta_0 = o_{\mathbb{P}^*}(1).$$

(ii) Assume **(A2)**, **(A3)**, **(A4)** and **(AB1)**–**(AB3)**. Then

$$\theta_m - \theta_0 = o_{\mathbb{P}_w^*}(1) \text{ i.p.}$$

Note that, the result of part (i) holds if we replaced **(A2)** and **(A4)** by **((A1')–(A3'))** and the result of part (ii) holds if we replaced **(A2)** and **(A4)** by **(A1')**, **(A2')** and **(AB1')**.

In the sequel, we refer to the sets of assumptions which implies the parts (i) and (ii); **(C)** and **(CP)**; respectively. Next we give the set of assumptions needed to identify rates of convergence of θ_n and θ_m to θ_0 , which is the important step for studying the weak convergence of these estimators.

Remark 3.3 We highlight that the parameter of interest θ is not restricted to belong to some Euclidean space as in Delsol and Van Keilegom (2020). More precisely, we consider the general framework in which $\theta \in \Theta$, where Θ is a subset of some Banach space \mathcal{B} . Notice that the result (i) of Theorem 3.2 is a bit more general than the analogous stated in the last reference, by the fact the conditions imposed are more general in our setting and extend those of Lee (2012) to the semiparametric models.

3.2 Rates of Convergence

Let us introduce the following assumptions:

(B1) $v_n d_{\mathcal{H}}(\hat{h}, h^0) = O_{\mathbb{P}^*}(1)$ for some $v_n \rightarrow \infty$.

(B2) For all $\delta_1 > 0$, there exist $\alpha < \nu$, $K > 0$, $\delta_0 > 0$, for all $n \in \mathbb{N}$ there exist a function φ for which $\delta \mapsto \frac{\varphi(\delta)}{\delta^\alpha}$ is decreasing on $(0, \delta_0]$ and $r_n^\nu \ell(r_n) n^{-1/2} \varphi(1/r_n) \leq C$ for n sufficiently large and some positive constant C , such that for all $\delta \leq \delta_0$,

$$\mathbb{P}^* \left[\sup_{\|\theta - \theta_0\| \leq \delta, d_{\mathcal{H}}(h, h^0) \leq \frac{\delta_1}{v_n}} |\mathbb{G}_n \tilde{\mathbf{m}}_{\theta - \theta_0, h}| \right] \leq K \varphi(\delta).$$

(B3) There exist $\eta_0 > 0$, $C > 0$ and two positive and non-decreasing functions ψ_1 and ψ_2 on $(0, \eta_0]$ such that for all $\theta \in \Theta$ satisfying $\|\theta - \theta_0\| \leq \delta_0$:

$$\mathbb{P} \tilde{\mathbf{m}}_{\theta - \theta_0, \hat{h}} \leq W_n \psi_1(\|\theta - \theta_0\|) - (C + o_{\mathbb{P}^*}(1)) \psi_2(\|\theta - \theta_0\|).$$

Moreover, there exist $\beta_2 > \alpha$, $\beta_1 < \beta_2$, $\delta_0 > 0$ such that $\delta \mapsto \psi_1(\delta) \delta^{-\beta_1}$ is non-increasing and $\delta \mapsto \psi_2(\delta) \delta^{-\beta_2}$ is non-decreasing on $(0, \delta_0]$, and such that, for some sequence $r_n \rightarrow \infty$,

$$\psi_1(r_n^{1-\nu} \ell^{-1}(r_n)) W_n = O_{\mathbb{P}^*}(\psi_2(r_n^{1-\nu} \ell^{-1}(r_n))).$$

for definition of \mathbb{P} -measurability.

(BB1) $v_m d_{\mathcal{H}}(\hat{h}_m, h^0) = O_{\mathbb{P}_w^*}(1)$ i.p. for some $v_m \rightarrow \infty$.

(BB2) With the same notation in assumption **(B2)** we replace $r_n (v_n)$ by $r_m (v_m)$ with assumption **(AB1)** we have;

$$\mathbb{P}^* \mathbb{P}_W^* \left[\sup_{\|\theta - \theta_0\| \leq \delta, d_{\mathcal{H}}(h, h^0) \leq \frac{\delta_1}{v_m}} |\widehat{\mathbb{G}}_m \widetilde{\mathbf{m}}_{\theta - \theta_0, h}| \right] \leq K \varphi(\delta).$$

(BB3) With the same notation in assumption **(B3)** we replace r_n by r_m with assumption **(AB1)** in mind we have;

$$\mathbb{P} \widetilde{\mathbf{m}}_{\theta - \theta_0, \widehat{h}_m} \leq W_m \boldsymbol{\psi}_1(\|\theta - \theta_0\|) - (C + o_{\mathbb{P}^*}(1)) \boldsymbol{\psi}_2(\|\theta - \theta_0\|),$$

where for some sequence $r_m \rightarrow \infty$,

$$\boldsymbol{\psi}_1(r_m^{1-\nu} \ell^{-1}(r_m)) W_m = O_{\mathbb{P}^*}(\boldsymbol{\psi}_2(r_m^{1-\nu} \ell^{-1}(r_m))), \text{ i.p.}$$

Remark 3.4

- (i) Assumption **(B1)** is a high-level assumption. Such condition on the nuisance parameter \widehat{h} could be obtained by many asymptotic results. In the present paper, we are primarily concerned with the cases where the convergence rate of the M -estimator of θ is not affected by the estimation of the nuisance parameter h .
- (ii) Assumption **(B2)** is fulfilled if we assume that for any \mathbf{x} the function $(\theta, h) \rightarrow \mathbf{m}(\mathbf{x}, \theta, h(\mathbf{x}, \theta)) - \mathbf{m}(\mathbf{x}, \theta_0, h(\mathbf{x}, \theta_0))$ is uniformly bounded on an open neighborhood of (θ_0, h^0) , i.e., on

$$\{(\theta, h) : \|\theta - \theta_0\| \leq \delta_0, d_{\mathcal{H}}(h, h^0) \leq \delta'_1\},$$

for some $\delta_0, \delta'_1 > 0$. We consider the class $\mathcal{M}_{\delta, \delta'_1}$ for any $\delta \leq \delta_0$ and its envelope M_{δ, δ'_1} . For any δ_1 , we have, for n large enough; $\delta_1 v_n^{-1} \leq \delta'_1$. After by the entropy conditions on $\mathcal{M}_{\delta, \delta'_1}$,

$$\int_0^1 \sup_{\delta < \delta_0} \sup_{\mathbb{Q}} \sqrt{1 + \log N \left(\epsilon \left\| M_{\delta, \delta'_1} \right\|_{\mathbb{L}_2(\mathbb{Q})}, \mathcal{M}_{\delta, \delta'_1}, \mathbb{L}_2(\mathbb{Q}) \right)} d\epsilon < +\infty, \tag{4}$$

where the second supremum is taken over all finitely discrete probability measures \mathbb{Q} on \mathcal{S} . We get;

$$\mathbb{P}^* \left[\sup_{\|\theta - \theta_0\| \leq \delta, d_{\mathcal{H}}(h, h^0) \leq \frac{\delta_1}{v_n}} |\widehat{\mathbb{G}}_n \widetilde{\mathbf{m}}_{\theta - \theta_0, h}| \right] \leq K_1 \sqrt{\mathbb{P}^* \left[M_{\delta, \delta'_1}^2 \right]},$$

see Theorems 2.14.1 and 2.14.2 in van der Vaart and Wellner (1996). Then the last part of **(B2)** holds if $\varphi(\delta)$ can be chosen such that

$$\exists K_0, \forall \delta \leq \delta_0 : \sqrt{\mathbb{P}^* \left[M_{\delta, \delta'_1}^2 \right]} \leq K_0 \varphi(\delta). \tag{5}$$

Note that, all the different rate of convergence r_n in the literature for smooth or not smooth function satisfied the last term in assumption **(B2)**.

(iii) We choose for simplification $\boldsymbol{\psi}_1(x) = Id(x) = x$ and $\boldsymbol{\psi}_2(x) = x^2$ in assumption **(B3)**, so it's hold under the following conditions :

- (a) $\Theta \subset \mathcal{B}$, where \mathcal{B} is a Banach space.
- (b) There exists $\delta_2 > 0$ such that for any h satisfying $d_{\mathcal{H}}(h, h^0) \leq \delta_2$, the function $\theta \mapsto \mathbb{P}(\mathbf{m}(\mathbf{X}, \theta, h))$ is twice Fréchet differentiable on an open neighborhood of θ_0 ,

$$\lim_{\|\theta - \theta_0\| \rightarrow 0} \sup_{d_{\mathcal{H}}(h, h^0) \leq \delta_2} \left\| \mathbb{P}\mathbf{m}_{\theta, h} - \mathbb{P}\mathbf{m}_{\theta_0, h} - \Gamma(\theta_0, h)(\theta - \theta_0) + \frac{1}{2} \Lambda(\theta_0, h)(\theta - \theta_0, \theta - \theta_0) \right\| = 0.$$

For more detail see (Allaire 2005, p.306).

- (c) $\Gamma(\theta_0, h)(\cdot)$ is a continuous linear form, with $\|\Gamma(\theta_0, \hat{h})\| = O_{\mathbb{P}^*} \left(\frac{1}{r_n^{v-1} \ell(r_n)} \right)$ and $\Gamma(\theta_0, h^0) = 0$.
- (d) $\Lambda(\theta_0, h)(\cdot, \cdot)$ is bilinear form with $\Lambda(\theta_0, h^0)$ is bounded, symmetric, positive definite and elliptic (i.e. $\Lambda(\theta_0, h^0)(u, u) \geq C\|u\|^2$) and $h \mapsto \Lambda(\theta_0, h)$ is continuous in h^0 , i.e.,

$$\lim_{d_{\mathcal{H}}(h, h^0) \rightarrow 0} \sup_{u \in \mathbb{R}^k, \|u\|=1} \left\| (\Lambda(\theta_0, h) - \Lambda(\theta_0, h^0))u \right\| = 0.$$

These assumptions and the fact that the bilinear form is bounded, it results when $d_{\mathcal{H}}(\hat{h}, h^0) \leq \delta_2$;

$$\begin{aligned} \mathbb{P}\mathbf{m}_{\theta, \hat{h}} - \mathbb{P}\mathbf{m}_{\theta_0, \hat{h}} &= \Gamma(\theta_0, \hat{h})(\gamma_\theta) - \frac{1}{2} \Lambda(\theta_0, h^0)(\gamma_\theta, \gamma_\theta) + o_{\mathbb{P}^*}(\|\gamma_\theta\|^2) + o(\|\gamma_\theta\|^2) \\ &\leq W_n \|\gamma_\theta\| - C \|\gamma_\theta\|^2 + \beta_n(\|\gamma_\theta\|), \end{aligned}$$

where $\gamma_\theta = \theta - \theta_0$. So **(B3)** holds with

$$W_n = \left\| \Gamma(\theta_0, \hat{h}) \right\|.$$

Note that when the space $\Theta \subset \mathbf{E}$ where \mathbf{E} is some Euclidean space, the Fréchet derivatives $\Gamma(\theta_0, h)$ and $\Lambda(\theta_0, h)$ become the usually derivatives i.e., the Gradient and Hessian matrix respectively, which is given in **Remark 2(v)** of Delsol and Van Keilegom (2020).

- (iv) Assumption **(BB1)** poses no difficulty in practice and is met trivially by, for example, setting $\hat{h}_m = \hat{h}$, like in **Remark 3.1** (iv).
- (v) Assumption **(BB2)** is a 'high-level' assumption. It serves to control the modulus of continuity of the bootstrapped empirical processes; which is needed to derive the rate of convergence of the bootstrapped estimator θ_m . Note that when we are in the situation of the n out of n bootstrap this condition is given in Ma and Kosorok (2005) and in Lemma 1 of Cheng and Huang (2010) for more generally in the exchangeable bootstrap weights. In our setting; it's fulfilled under some entropy conditions, for brevity with the same notation in **(ii)**, let $\tilde{N}_1, \tilde{N}_2, \dots$ be i.i.d. symmetrized Poisson variables with parameter $\frac{1}{2}m/n$ and $\varepsilon_1, \varepsilon_2, \dots$ are i.i.d. Rademacher variables inde-

pendent of $\tilde{N}_1, \tilde{N}_2, \dots$ and $\mathbf{X}_1, \mathbf{X}_2, \dots$. Denote by $R = (R_1, \dots, R_n)$ a random permutation of $\{1, 2, \dots, n\}$, independent of all other variables. Let us introduce

$$\mathbb{P}_k^R = k^{-1} \sum_{i=1}^k \delta_{\mathbf{X}_{R_i}},$$

for each $k \in \{1, \dots, n\}$. By Lemma 3.6.6 of van der Vaart and Wellner (1996) and the paragraph before it (ahead) with sub-Gaussian inequality for Rademacher process we obtain

$$\mathbb{P}_W^* \left\| \hat{\mathbb{G}}_m \right\|_{\mathcal{M}_{\delta, \delta'_1}} \leq 4 \mathbb{P}_{\tilde{N}} \left\| \frac{1}{\sqrt{k}} \sum_{i=1}^n |\tilde{N}_i| \varepsilon_i \delta_{\mathbf{X}_i} \right\|_{\mathcal{M}_{\delta, \delta'_1}}. \tag{6}$$

Applying now Lemma 3.6.7 of van der Vaart and Wellner (1996) to the right side of (6) with n_0 set to 1 we get;

$$\begin{aligned} \mathbb{P}_W^* \left\| \hat{\mathbb{G}}_m \right\|_{\mathcal{M}_{\delta, \delta'_1}} &\leq 4 \mathbb{P}_{\tilde{N}} \left\| \frac{1}{\sqrt{k}} \sum_{i=1}^n |\tilde{N}_i| \varepsilon_i \delta_{\mathbf{X}_i} \right\|_{\mathcal{M}_{\delta, \delta'_1}} \\ &\leq \sqrt{\frac{n}{k}} \left\| \tilde{N}_i \right\|_{2,1} \max_{1 \leq k \leq n} \mathbb{P}_R \mathbb{P}_\varepsilon \left\| \frac{1}{\sqrt{k}} \sum_{i=1}^k \varepsilon_i \delta_{\mathbf{X}_{R_i}} \right\|_{\mathcal{M}_{\delta, \delta'_1}}^* \\ &\leq C \max_{1 \leq k \leq n} \mathbb{P}_R \left(\mathbb{P}_k^R M_{\delta, \delta'_1} \right)^{1/2} \\ &\leq C \left(\mathbb{P}_n M_{\delta, \delta'_1} \right)^{1/2}, \end{aligned} \tag{7}$$

where $C > \sqrt{\frac{n}{k}} \left\| \tilde{N}_i \right\|_{2,1}$ see Problem 3.6.3 of van der Vaart and Wellner (1996). By Jensen’s inequality the outer expectation of the left side of (7) is bounded by $C \sqrt{\mathbb{P}[M_{\delta, \delta'_1}]^2}$, for every $\delta < \delta_1$. The inequality in assumption (BB2) holds for every $n \in \mathbb{N}$ this implied by the fact that the variables we consider are i.i.d.

- (vi) Finally, for the assumption (BB3) with the same discussion given in (iii) only the choice $W_n = \left\| \Gamma(\theta_0, \hat{h}) \right\|$ becomes

$$W_m = \left\| \Gamma(\theta_0, \hat{h}_m) \right\|,$$

with $W_m = O_{\mathbb{P}_W^*} \left(\frac{1}{r_m^{s-1} \ell(r_m)} \right)$ i.p.

Theorem 3.5

- (i) Assume (C) and (B1)-(B3). Then

$$r_n(\theta_n - \theta_0) = O_{\mathbb{P}^*}(1).$$

- (ii) Assume (CP) and (BB1)-(BB3). Then

$$r_m(\theta_m - \theta_0) = O_{\mathbb{P}_W^*}(1) \text{ i.p.}$$

Remark 3.6 The result (i) of this Theorem still holds for θ belongs to Banach space which is more general of the Theorem 2 of Delsol and Van Keilegom (2020), where the authors are interested in the finite dimensional case. Noting that the choice of $v = 2$ and $\ell \equiv 1$ in assumptions **B2** and **B3**, reduces to the assumptions **B2** and **B3** respectively of the last reference.

3.3 Weak Convergence

We start this section by introducing some notation. For any $\theta \in \Theta$ and $h \in \mathcal{H}$, let $\mathcal{K} = \{\gamma \in \mathbf{E} : \|\gamma\| \leq K\}$ for $K > 0$. Define, for sufficiently large n and for $\gamma \in \mathcal{K}$, the empirical processes

$$\begin{aligned} \mathbb{M}_n(\gamma, h) &= r_n^v \ell(r_n) (\mathbb{P}_n - \mathbb{P}) \tilde{\mathbf{m}}_{\gamma/r_n, h}, \\ \widehat{\mathbb{M}}_n(\gamma, h) &= r_m^v \ell(r_m) (\widehat{\mathbb{P}}_m - \mathbb{P}_n) \tilde{\mathbf{m}}_{\gamma/r_m, h}, \end{aligned} \tag{8}$$

which can be treated as random elements in $\ell^\infty(\mathcal{K})$. Also let for any $\delta > 0$;

$$\begin{aligned} \mathbf{M}_\delta(\cdot) &\geq \sup_{\|\theta - \theta_0\| \leq \delta} \left| \mathbf{m}(\cdot, \theta, h^0) - \mathbf{m}(\cdot, \theta_0, h^0) \right|, \\ \mathcal{M}_\delta &= \left\{ \mathbf{m}(\cdot, \theta, h^0) - \mathbf{m}(\cdot, \theta_0, h^0) : \|\theta - \theta_0\| \leq \delta \right\}, \\ \mathcal{M}_\delta(\eta) &= \left\{ \mathbf{m}_{\theta, h^0} - \mathbf{m}_{\psi, h^0} : \|\theta - \psi\| < \eta, \|\theta - \theta_0\| \vee \|\psi - \theta_0\| < \delta, \theta, \psi \in \Theta \right\}. \end{aligned}$$

Finally, for any $p \in \mathbb{N}$ and any $f : \Theta \rightarrow \mathbb{R}$ and for any $\gamma = (\gamma_1, \dots, \gamma_p) \in \Theta^p$, denote

$$\bar{f}_\gamma = (f(\gamma_1), \dots, f(\gamma_p))^\top.$$

We give the set of assumptions for the asymptotic distribution of the processes given in (8) and their maximum.

- (C1) $r_n \|\theta_n - \theta_0\| = O_{\mathbb{P}^*}(1)$ and $v_n d_{\gamma_t}(\hat{h}, h^0) = O_{\mathbb{P}^*}(1)$ for some sequences $r_n \rightarrow \infty$ and $v_n \rightarrow \infty$, and $r_n^{v-2} \ell(r_n) < C$ for some $C > 0$.
- (C2) θ_0 lies to the interior of Θ , where $\Theta \subset (\mathcal{B}, \|\cdot\|)$.
- (C3) For all $\delta_2, \delta_3 > 0$,

$$\begin{aligned} \sup_{\|\theta - \theta_0\| \leq \frac{\delta_2}{r_n}} \frac{|\mathbb{P}_n - \mathbb{P} \tilde{\mathbf{m}}_{\theta - \theta_0, h} + (\mathbb{P}_n - \mathbb{P}) \tilde{\mathbf{m}}_{\theta - \theta_0, h^0}|}{r_n^{-v} \ell^{-1}(r_n) + |\mathbb{P}_n \tilde{\mathbf{m}}_{\theta - \theta_0, h}| + |\mathbb{P}_n \tilde{\mathbf{m}}_{\theta - \theta_0, h^0}| + |\mathbb{P} \tilde{\mathbf{m}}_{\theta - \theta_0, h}| + |\mathbb{P} \tilde{\mathbf{m}}_{\theta - \theta_0, h^0}|} &= o_{\mathbb{P}^*}(1), \\ d_{\gamma_t}(h, h^0) \leq \frac{\delta_3}{v_n} \end{aligned}$$

- (C4) There exists a sequence $\{a_n\}$ with

$$a_m^2 m^{-1/2} \log n / \log(n/m + 1) = o(1) \quad \text{and} \quad a_n^{-1} = O(1),$$

such that, for all $C, \eta > 0$ and for every sequence $\{j_n\}$ with $a_n = o(j_n)$,

$$\frac{r_n^{2v} \ell^2(r_n)}{n} \mathbb{P}^* \left[\mathbf{M}_{\frac{C}{r_n}}^2 \right] = O(1) \quad \text{and} \quad \frac{r_n^{2v} \ell^2(r_n)}{n} \mathbb{P}^* \left[\mathbf{M}_{\frac{C}{r_n}}^2 \mathbb{I} \left\{ \mathbf{M}_{\frac{C}{r_n}} > \frac{\eta j_n^{1/2}}{r_n^v \ell(r_n)} \right\} \right] = o(1).$$

- (C5) For all K and for any $\eta_n \rightarrow 0$,

$$\sup_{\|\gamma_1 - \gamma_2\| < \eta_n, \|\gamma_1\| \vee \|\gamma_2\| \leq K} \frac{r_n^{2\nu} \ell^2(r_n)}{n} \mathbb{P} \left[\mathbf{m} \left(\mathbf{X}, \boldsymbol{\theta}_0 + \frac{\boldsymbol{\gamma}_1}{r_n}, h^0 \right) - \mathbf{m} \left(\mathbf{X}, \boldsymbol{\theta}_0 + \frac{\boldsymbol{\gamma}_2}{r_n}, h^0 \right) \right]^2 = o(1).$$

(C6) For \mathbf{x} , the function $\boldsymbol{\theta} \mapsto \mathbf{m}(\mathbf{x}, \boldsymbol{\theta}, h^0)$ and almost all paths of the two processes $\boldsymbol{\theta} \mapsto \mathbf{m}(\mathbf{x}, \boldsymbol{\theta}, \hat{h})$ and $\boldsymbol{\theta} \mapsto \mathbf{m}(\mathbf{x}, \boldsymbol{\theta}, \hat{h}_m)$ are uniformly bounded on closed bounded sets (over $\boldsymbol{\theta}$).

(C7) There exist a random and linear function $W_n : \mathcal{B} \rightarrow \mathbb{R}$, a deterministic and bilinear function $V : \mathcal{B} \times \mathcal{B} \rightarrow \mathbb{R}$ and $\beta_n = o_{\mathbb{P}^*}(1)$; such that for all $\boldsymbol{\theta} \in \Theta$;

$$\mathbb{P} \tilde{\mathbf{m}}_{\boldsymbol{\theta} - \boldsymbol{\theta}_0, \hat{h}} = W_n(\boldsymbol{\gamma}_\boldsymbol{\theta}) + V(\boldsymbol{\gamma}_\boldsymbol{\theta}, \boldsymbol{\gamma}_\boldsymbol{\theta}) + \beta_n \|\boldsymbol{\gamma}_\boldsymbol{\theta}\|^2 + o(\|\boldsymbol{\gamma}_\boldsymbol{\theta}\|^2)$$

and

$$\mathbb{P} \tilde{\mathbf{m}}_{\boldsymbol{\theta} - \boldsymbol{\theta}_0, h^0} = V(\boldsymbol{\gamma}_\boldsymbol{\theta}, \boldsymbol{\gamma}_\boldsymbol{\theta}) + o(\|\boldsymbol{\gamma}_\boldsymbol{\theta}\|^2),$$

where $\boldsymbol{\gamma}_\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_0$ and the notation $o(\|\boldsymbol{\gamma}_\boldsymbol{\theta}\|^2)$ means

$$\lim_{\|\boldsymbol{\gamma}_\boldsymbol{\theta}\| \rightarrow 0} \frac{o(\|\boldsymbol{\gamma}_\boldsymbol{\theta}\|^2)}{\|\boldsymbol{\gamma}_\boldsymbol{\theta}\|^2} = 0.$$

Moreover, for any bounded closed set $\mathcal{K} \subset \mathcal{B}$,

$$\exists \tau, \delta_1 > 0, r_n^{\nu-1} \ell(r_n) \sup_{\substack{\boldsymbol{\gamma} \in \mathcal{K}, \delta \leq \delta_1 \\ \|\boldsymbol{\gamma}\| \leq \delta}} \left| \frac{W_n(\boldsymbol{\gamma})}{\delta^\tau} \right| = O_{\mathbb{P}}(1) \quad \text{and} \quad \sup_{\substack{\boldsymbol{\gamma}, \boldsymbol{\gamma}' \in \mathcal{K}, \delta \leq \delta_1 \\ \|\boldsymbol{\gamma} - \boldsymbol{\gamma}'\| \leq \delta}} \frac{|V(\boldsymbol{\gamma}, \boldsymbol{\gamma}) - V(\boldsymbol{\gamma}', \boldsymbol{\gamma}')|}{\delta^\tau} < \infty.$$

(C8) There exists a zero-mean Gaussian process \mathbb{G} defined on \mathcal{B} and a continuous function Λ such that for all $p \in \mathbb{N}$ and for all $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_p) \in \mathcal{K}^p$,

$$r_n^{\nu-1} \ell(r_n) \overline{W_{n\boldsymbol{\gamma}}} + r_n^\nu \ell(r_n) \overline{\mathbb{P}_n \tilde{\mathbf{m}}_{\boldsymbol{\gamma}, h^0}} \Rightarrow \overline{\Lambda}_\boldsymbol{\gamma} + \overline{\mathbb{G}}_\boldsymbol{\gamma}.$$

Moreover, $\mathbb{G}(\boldsymbol{\gamma}) = \mathbb{G}(\boldsymbol{\gamma}')$ a.s. implies that $\boldsymbol{\gamma} = \boldsymbol{\gamma}'$, and

$$\mathbb{P}^* \left(\limsup_{\|\boldsymbol{\gamma}\| \rightarrow \infty} (\Lambda(\boldsymbol{\gamma}) + \mathbb{G}(\boldsymbol{\gamma})) < \sup_{\boldsymbol{\gamma} \in \mathcal{B}} (\Lambda(\boldsymbol{\gamma}) + \mathbb{G}(\boldsymbol{\gamma})) \right) = 1.$$

(C9) There exists a $\delta_0 > 0$ such that

$$\int_0^\infty \sup_{\delta \leq \delta_0} \sup_{\mathbb{Q}} \sqrt{\log (N(\epsilon \|\mathbf{M}_\delta\|_{\mathbb{Q}, 2}, \mathcal{M}_\delta, \mathbb{L}^2(\mathbb{Q})))} d\epsilon < \infty.$$

(C10) For all $\delta, \eta > 0$, the classes $\mathcal{M}_\delta, \mathcal{M}_\delta(\eta)$ and $\mathcal{M}_\delta(\eta)^2$ are \mathbb{P} -measurable, see (van der Vaart and Wellner 1996, p.110) for definition of \mathbb{P} -measurability.

(C11) For all $C > 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n_0 \geq n$,

$$\mathbb{P}_n \mathbf{m}_{\boldsymbol{\theta}_n, \hat{h}} \geq \sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \frac{C}{r_n}} \mathbb{P}_n \mathbf{m}_{\boldsymbol{\theta}_0, \hat{h}} - R_n,$$

where R_n is given in (A5).

(CB1) $r_m \|\theta_m - \theta_0\| = O_{\mathbb{P}_w^*}(1)$ i.p. and $v_m d_{\mathcal{H}}(\hat{h}_m, h^0) = O_{\mathbb{P}_w^*}(1)$ i.p. for some sequences $r_m \rightarrow \infty$ and $v_m \rightarrow \infty$ and $r_m^{\nu-2} \ell(r_m) \leq C$.

(CB2) For all $\delta_2, \delta_3 > 0$,

$$\sup_{\|\theta - \theta_0\| \leq \frac{\delta_2}{r_m}, d_{\mathcal{H}}(h, h^0) \leq \frac{\delta_3}{v_m}} \frac{|(\hat{\mathbb{P}}_m - \mathbb{P}_n)\tilde{\mathbf{m}}_{\theta-\theta_0, h} + (\hat{\mathbb{P}}_m - \mathbb{P}_n)\tilde{\mathbf{m}}_{\theta-\theta_0, h^0}|}{r_m^{-\nu} \ell^{-1}(r_m) + |\mathbb{P}_n \tilde{\mathbf{m}}_{\theta-\theta_0, h}| + |\mathbb{P}_n \tilde{\mathbf{m}}_{\theta-\theta_0, h^0}| + |\hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{\theta-\theta_0, h}| + |\hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{\theta-\theta_0, h^0}|} = o_{\mathbb{P}^*}(1).$$

(CB3) There exists a random and linear function $W_m : \mathcal{B} \rightarrow \mathbb{R}$, and $\beta_m = o_{\mathbb{P}^*}(1)$, such that for all $\theta \in \Theta$;

$$\mathbb{P} \tilde{\mathbf{m}}_{\theta-\theta_0, \hat{h}_m} = W_m(\boldsymbol{\gamma}_\theta) + V(\boldsymbol{\gamma}_\theta, \boldsymbol{\gamma}_\theta) + \beta_m \|\boldsymbol{\gamma}_\theta\|^2 + o(\|\boldsymbol{\gamma}_\theta\|^2)$$

and

$$\mathbb{P} \tilde{\mathbf{m}}_{\theta-\theta_0, h^0} = V(\boldsymbol{\gamma}_\theta, \boldsymbol{\gamma}_\theta) + o(\|\boldsymbol{\gamma}_\theta\|^2).$$

Moreover, for any closed bounded set $\mathcal{K} \subset \mathbf{E}$,

$$\exists \tau, \delta_1 > 0, r_m^{\nu-1} \ell(r_m) \sup_{\substack{\boldsymbol{\gamma} \in \mathcal{K}, \delta \leq \delta_1 \\ \|\boldsymbol{\gamma}\| \leq \delta}} \left| \frac{W_m(\boldsymbol{\gamma})}{\delta^\tau} \right| = O_{\mathbb{P}_w^*}(1) \text{ i.p.}, \sup_{\substack{\boldsymbol{\gamma}, \boldsymbol{\gamma}' \in \mathcal{K}, \delta \leq \delta_1 \\ \|\boldsymbol{\gamma} - \boldsymbol{\gamma}'\| \leq \delta}} \frac{|V(\boldsymbol{\gamma}, \boldsymbol{\gamma}) - V(\boldsymbol{\gamma}', \boldsymbol{\gamma}')|}{\delta^\tau} < \infty.$$

(CB4)

$$r_m^{\nu-1} \ell(r_m) \overline{W_m \boldsymbol{\gamma}} + r_m^\nu \ell(r_m) \overline{\hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{\cdot, h^0} \boldsymbol{\gamma}} \Rightarrow \overline{\Lambda} \boldsymbol{\gamma} + \overline{\mathbb{G}} \boldsymbol{\gamma} \text{ i.p.},$$

where Λ and \mathbb{G} are given in **(C8)** and the weak convergence is conditionally on the sample.

(CB5) For all $C > 0$, there exist $m_0 \in \mathbb{N}$ such that for all $m \geq m_0$,

$$\hat{\mathbb{P}}_m \mathbf{m}_{\theta_m, \hat{h}_m} \geq \sup_{\|\theta - \theta_0\| \leq \frac{C}{r_m}} \hat{\mathbb{P}}_m \mathbf{m}_{\theta_0, \hat{h}_m} - \hat{R}_n,$$

where \hat{R}_n is given in **(AB3)**.

Remark 3.7

- (i) We can obtained the first part of condition **(C1)** by part (i) of Theorem 3.5.
- (ii) Assumption **(C3)** holds under the common condition: for all $\delta_2, \delta_3 > 0$,

$$\sup_{\|\theta - \theta_0\| \leq \frac{\delta_2}{r_n}, d_{\mathcal{H}}(h, h^0) \leq \frac{\delta_3}{v_n}} \left| (\mathbb{P}_n - \mathbb{P}) \tilde{\mathbf{m}}_{\theta-\theta_0, h} + (\mathbb{P}_n - \mathbb{P}) \tilde{\mathbf{m}}_{\theta-\theta_0, h^0} \right| = o_{\mathbb{P}^*}(r_n^{-\nu} \ell^{-1}(r_n)),$$

which is implied by the fact that; there exists a function f and a constant $\delta_0 > 0$ such that for all $\delta_2, \delta_3 < \delta_0$,

$$r_n^v \ell(r_n) f\left(\frac{\delta_2}{r_n}, \frac{\delta_3}{v_n}\right) = o(\sqrt{n}),$$

and

$$\begin{aligned} & \mathbb{P}^* \left[\sup_{\|\theta - \theta_0\| \leq \frac{\delta_2}{r_n}, d_{\mathcal{H}}(h, h^0) \leq \frac{\delta_3}{v_n}} \left| (\mathbb{P}_n - \mathbb{P}) \tilde{\mathbf{m}}_{\theta - \theta_0, h} + (\mathbb{P}_n - \mathbb{P}) \tilde{\mathbf{m}}_{\theta - \theta_0, h^0} \right| \right] \\ & \leq 2 \mathbb{P}^* \left[\sup_{\|\theta - \theta_0\| \leq \frac{\delta_2}{r_n}, d_{\mathcal{H}}(h, h^0) \leq \frac{\delta_3}{v_n}} \left| \mathbb{G}_n \tilde{\mathbf{m}}_{\theta - \theta_0, h} \right| \right] \\ & \leq \frac{1}{\sqrt{n}} f\left(\frac{\delta_2}{r_n}, \frac{\delta_3}{v_n}\right). \end{aligned}$$

Using the same arguments as in **Remark 3.3(ii)**, we get the last inequality.

- (iii) If we assume that $j_n = \sqrt{n}$, and noting; $\gamma \mapsto \mathbb{M}_n(\gamma, h^0) = \frac{r_n^v \ell(r_n)}{\sqrt{n}} \mathbb{G}_n \tilde{\mathbf{m}}_{\gamma/r_n, h^0}$ is the empirical process with indexed class $\frac{r_n^v \ell(r_n)}{\sqrt{n}} \mathcal{M}_{\frac{c}{r_n}}$ then, under assumption **(B2)**, the assumptions **(C4)** and **(C5)** hold by the following conditions: there exists a $\delta_4 > 0$ such that for all $\delta \leq \delta_4$, $\mathbb{P}^*(\mathbf{M}_\delta^2) \leq K\varphi^2(\delta)$ for some $C > 0$,

$$\lim_{\delta \rightarrow 0} \frac{\mathbb{P}^* [\mathbf{M}_\delta^2 \mathbb{I}_{\{\mathbf{M}_\delta > \eta \delta^{-2} \varphi^2(\delta)\}}]}{\varphi^2(\delta)} = 0,$$

for all $\eta > 0$ and

$$\lim_{\epsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \sup_{\|\gamma_1 - \gamma_2\| < \epsilon, \|\gamma_1\| \vee \|\gamma_2\| \leq K} \frac{\mathbb{P} [\mathbf{m}(\mathbf{X}, \theta_0 + \gamma_1 \delta, h^0) - \mathbf{m}(\mathbf{X}, \theta_0 + \gamma_2 \delta, h^0)]^2}{\varphi^2(\delta)} = 0,$$

for all $C > 0$, corresponding the case to Theorem 3.2.10 in van der Vaart and Wellner (1996).

- (iv) Let \mathcal{K} be an arbitrary closed bounded subset in \mathcal{B} , the first part of condition **(C8)** is used to assume the convergence of the marginals of the process $\gamma \mapsto r_n^{v-1} \ell(r_n) W_n(\gamma) + r_n^v \ell(r_n) \mathbb{P}_n \tilde{\mathbf{m}}_{\frac{\gamma}{r_n}, h^0}$ for deriving its weak convergence in $\ell^\infty(\mathcal{K})$ by the fact that it is asymptotically tight; which is fulfilling by using **(C4)**, **(C5)**, **(C9)** and the preceding discussion in **(iii)**. If

$$r_n^{v-1} \ell(r_n) \sup_{\gamma \in \mathcal{K}, \gamma \neq 0} \|W_n(\gamma)\| \|\gamma\|^{-1} = o_{\mathbb{P}}(1),$$

we treat the given process as in the parametric case, where its marginals converge provided that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{r_n^{2v} \ell^2(r_n)}{n} \mathbb{P} \left\{ \left[\mathbf{m}\left(\mathbf{X}, \theta_0 + \frac{\gamma_1}{r_n}, h^0\right) - \mathbf{m}\left(\mathbf{X}, \theta_0 + \frac{\gamma_2}{r_n}, h^0\right) \right]^2 \right\} \\ & = \mathbb{P} \left[(\mathbb{G}(\gamma_1) - \mathbb{G}(\gamma_2))^2 \right], \end{aligned}$$

for all γ_1, γ_2 and we lead to a rate of convergence r_n as the solution of

$$r_n^\nu \ell(r_n) \varphi(1/r_n) = \sqrt{n},$$

for more detail see Theorem 3.2.10 of van der Vaart and Wellner (1996). Note that almost all sample paths of the process $\gamma \mapsto \Lambda(\gamma) + \mathbb{G}(\gamma)$ have a supremum affiliated to their attitude on bounded closed set, which is guaranteed by the last assumption. The dominant term of the deterministic part Λ is usually a negative definite quadratic form and hence exponential inequalities could lead to such result, for example when we are in the situation of the smooth function, one can refer to Lee and Pun (2006), Ma and Kosorok (2005), Kosorok (2008), Kristensen and Salanié (2017) among many others.

- (v) Assumption (C9) is a technical assumption, which is the same in the parametric case where the nuisance parameter h^0 is known, needs to show that; the process $\gamma \mapsto r_n^\nu \ell(r_n) \mathbb{P}_n \tilde{\mathbf{m}}_{\theta-\theta_0, h^0}$ is asymptotically tight, see Theorem 3.2.10 of van der Vaart and Wellner (1996).
- (vi) First part of (CB1) follows by part (ii) of Theorem 3.5.
- (vii) Assumption (CB2) is automatically hold under the condition : for all $\delta_2, \delta_3 > 0$,

$$\sup_{\|\theta-\theta_0\| \leq \frac{\delta_2}{r_m}, d_{\mathcal{H}}(h, h^0) \leq \frac{\delta_3}{v_m}} \left| (\hat{\mathbb{P}}_m - \mathbb{P}_n) \tilde{\mathbf{m}}_{\theta-\theta_0, h} + (\hat{\mathbb{P}}_m - \mathbb{P}_n) \tilde{\mathbf{m}}_{\theta-\theta_0, h^0} \right| = o_{\mathbb{P}_W^*} (r_m^{-\nu} \ell^{-1}(r_m)) \text{ i.p.}$$

This condition is hold if: there exists a function g and a constant $\delta_0 > 0$ such that for all $\delta_2, \delta_3 < \delta_0$,

$$r_m^\nu \ell(r_m) g\left(\frac{\delta_2}{r_m}, \frac{\delta_3}{v_m}\right) = o(\sqrt{m}),$$

and

$$\begin{aligned} & \mathbb{P}_W^* \mathbb{P}_W^* \left[\sup_{\|\theta-\theta_0\| \leq \frac{\delta_2}{r_m}, d_{\mathcal{H}}(h, h^0) \leq \frac{\delta_3}{v_m}} \left| (\hat{\mathbb{P}}_m - \mathbb{P}_n) \tilde{\mathbf{m}}_{\theta-\theta_0, h} + (\hat{\mathbb{P}}_m - \mathbb{P}_n) \tilde{\mathbf{m}}_{\theta-\theta_0, h^0} \right| \right] \\ & \leq 2 \mathbb{P}_W^* \mathbb{P}_W^* \left[\sup_{\|\theta-\theta_0\| \leq \frac{\delta_2}{r_m}, d_{\mathcal{H}}(h, h^0) \leq \frac{\delta_3}{v_m}} \left| \hat{\mathbb{G}}_m \tilde{\mathbf{m}}_{\theta-\theta_0, h} \right| \right] \\ & \leq \frac{1}{\sqrt{m}} g\left(\frac{\delta_2}{r_m}, \frac{\delta_3}{v_m}\right). \end{aligned}$$

Using the same arguments as in Remark 3.3(v), we get the last inequality.

- (viii) Following similar discussion of the condition (C7) provided in Remark 3(iv) of Delsol and Van Keilegom (2020), we only change the random function $W_n(\gamma)$ for the bootstrap version to $W_m(\gamma) = \langle \Gamma(\theta_0, \hat{h}_m), \gamma \rangle$. If we are in the situation where \hat{h}_m is calculated from a dataset independently from the bootstrapped sample $(\mathbf{X}_1^*, \dots, \mathbf{X}_m^*)$, so it is sufficient for assumption (CB4) to suppose the conditional weak convergence of each term; $r_m^{-\nu-1} \ell(r_m) \overline{W}_{m\gamma}$ and $r_m^\nu \ell(r_m) \hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{\theta-\theta_0, h^0}$ separately. We can get the convergence of the second one as the same in the situation without the nuisance parameter, the interested reader is referred to Lemma 1 of Lee (2012). Note that if $r_m^{-\nu-1} \ell(r_m) \Gamma(\theta_0, \hat{h}_m) \rightarrow W$ conditionally in distribution, the marginals of the process

$\gamma \mapsto \langle r_m^{\nu-1} \ell(r_m) \Gamma(\theta_0, \hat{h}_m), \gamma \rangle$ tend in distribution to the marginals of $\gamma \mapsto \langle W, \gamma \rangle$. Furthermore, if $r_m = \sqrt{m}$ and $\ell \equiv 1$, it is common to assume that

$$\Gamma(\theta_0, \hat{h}_m) = m^{-1} \sum_{i=1}^m U_{i,m} + o_{\mathbb{P}_W^*}(m^{-1/2}),$$

where $U_{i,m}, i = 1, \dots, m$, are independent and centered random variables. The convergence follows from Lindeberg’s condition.

Theorem 3.8 (Weak Convergence of Empirical Processes). *For all $K > 0$, let $\mathcal{K} = \{\gamma \in \mathbf{E} : \|\gamma\| \leq K\}$ be a closed bounded subset of \mathcal{B} , treating $\gamma \mapsto \mathbb{M}_n(\gamma, \hat{h})$ and $\gamma \mapsto \hat{\mathbb{M}}_n(\gamma, \hat{h}_m)$ as random elements in $\ell(\mathcal{K})$ for sufficiently large n , we have the following results:*

- (i) Assume (C1)-(C10). Then

$$r_n^\nu \ell(r_n) \mathbb{P}_n \tilde{\mathbf{m}}_{r_n, \hat{h}} \Rightarrow \Lambda(\gamma) + \mathbb{G}(\gamma).$$

- (ii) Assume (A2), (AB1), (B2), (C2)-(C6), (C9)- (C11) and (CB1)-(CB4). Then

$$r_m^\nu \ell(r_m) \hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, \hat{h}_m} \Rightarrow \Lambda(\gamma) + \mathbb{G}(\gamma) \text{ i.p.}$$

Our main results concerning weak convergence of $r_n(\theta_n - \theta_0)$ and m out of n bootstrap consistency are embodied in the following theorem.

Theorem 3.9 *Assume for any such \mathcal{K} that almost every sample path of the process $\gamma \mapsto \Lambda(\gamma) + \mathbb{G}(\gamma)$ achieves its supremum at a unique random point $\gamma_0 = \arg \max_{\gamma \in \mathcal{B}} \Lambda(\gamma) + \mathbb{G}(\gamma)$, then;*

- (i) Assume (C1)-(C11). Then

$$r_n(\theta_n - \theta_0) \Rightarrow \gamma_0.$$

- (ii) Assume (A2), (AB1), (B2), the first part of (C1), (C2)-(C6), (C9)- (C11) and (CB1)-(CB5). Then

$$r_m(\theta_m - \theta_n) \Rightarrow \gamma_0 \text{ i.p.}$$

Remark 3.10 The result (i) of the Theorem 3.8 is the same result of Lemma 1 of Delsol and Van Keilegom (2020) where the parameter of the interest θ is in a Euclidean space, for the particular case $\nu = 2$ and $\ell \equiv 1$, then by the application of Theorem 3.2.2 of van der Vaart and Wellner (1996) and the uniform tightness of the sequence $r_n(\theta_n - \theta_0)$, the authors established the weak convergence to some tight random variable γ_0 in $\ell^\infty(\mathcal{K})$ for the compact set \mathcal{K} in their Theorem 3 which is given in the result (i) of the Theorem 3.9 in this case. In our setting, we provide the weak convergence of the same sequence for the Banach valued parameter by using Theorem of van der Vaart and Wellner (1996) where the compact sets and the uniform tightness of $r_n(\theta_n - \theta_0)$ are replaced, respectively, by closed

bounded sets with a similar structure as the set \mathcal{K} and $r_n(\theta_n - \theta_0) = O_{\mathbb{P}^*}(1)$, as given in Lee (2012) without the nuisance parameter h^0 .

Note that (i) still holds if (C4) is replaced by this more weak condition

$$n^{-1}r_n^{2\nu} \ell(r_n)^2 \mathbb{P}^* \mathbf{M}_{c/r_n}^2 \{ \mathbf{M}_{c/r_n} > \eta n r_n^{-\nu} \ell^{-1}(r_n) \} \rightarrow 0.$$

In order to prove the conditional stochastic equicontinuity of the bootstrapped process $\widehat{\mathbb{M}}_n$ we need the condition (C4), that is fulfilled if the uniform integrability condition is imposed for $j_n \geq n^c$, for some $0 < c < 1/4$.

Remark 3.11 It is well known that Theorem 3.9 can be used easily through routine bootstrap sampling, which we describe briefly as follows. More precisely, this can be used, for example, to form confidence bands for the true parameter θ based N , be a large integer, sampled samples $\mathbf{Y}_1^k, \dots, \mathbf{Y}_m^k, k = 1, \dots, \mathfrak{N}$. Let θ_m^k the bootstrapped estimator of θ based on the sample $\mathbf{Y}_1^k, \dots, \mathbf{Y}_m^k, k = 1, \dots, \mathfrak{N}$. An application of Theorem 3.9 implies that

$$(r_n(\theta_n - \theta_0), r_m(\theta_m^1 - \theta_n), \dots, r_m(\theta_m^{\mathfrak{N}} - \theta_n)) \Rightarrow (\gamma_0, \gamma_0^1, \dots, \gamma_0^{\mathfrak{N}}) \text{ i.p. ,}$$

where $\gamma_0^1, \dots, \gamma_0^{\mathfrak{N}}$ are independent copies of γ_0 . Notice that we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(\theta_n - r_n^{-1}c(\alpha) \leq \theta_0 \leq \theta_n + r_n^{-1}c(\alpha)) = \mathbb{P}(|\gamma_0| \leq c(\alpha)) = 1 - \alpha.$$

Table 1 Kolmogorov Distance (KD) Between Distributions of $n^{1/3}(\theta_n - \theta^0)$ and $m^{1/3}(\theta_m - \theta_n)$, for $n = 250, n = 1000$ and $n = 2000$

$n = 250$		$n = 1000$		$n = 2000$	
m	KD	m	KD	m	KD
10	0,1733	50	0,0880	50	0,0730
20	0,1267	60	0,0843	100	0,0780
30	0,0853	70	0,0870	110	0,0717
40	0,0687	80	0,1267	120	0,1150
50	0,0527	90	0,1197	130	0,0950
60	0,0793	100	0,1040	140	0,1103
70	0,0780	110	0,1180	150	0,0997
80	0,1213	120	0,1133	160	0,1437
90	0,0953	130	0,1080	170	0,1557
100	0,1183	140	0,1283	180	0,1337
125	0,1383	150	0,1073	190	0,1523
150	0,1453	200	0,1553	200	0,1480
175	0,1773	275	0,2187	300	0,2073
200	0,1757	350	0,2543	400	0,2530
225	0,2057	425	0,2917	500	0,2833
250	0,1993	500	0,2990	750	0,3533
		750	0,3787	1000	0,3953
		1000	0,4187	1250	0,4310
				1250	0,4310
				1500	0,4537
				2000	0,5240

In order to approximate $c(\alpha)$, one can use the sampling estimator $\hat{c}(\alpha)$, of $c(\alpha)$, as the smallest z such that

$$\frac{1}{\mathfrak{N}} \sum_{k=1}^{\mathfrak{N}} \mathbb{1}_{\{r_m \theta_m^k - \theta_n \leq z\}} \geq 1 - \alpha.$$

Remark 3.12 In nonregular problems where the conventional n out of n bootstrap is inconsistent, the m out of n bootstrap provides a useful remedy to restore consistency. In practice, however, choosing an appropriate m needs careful attention. Asymptotically, \sqrt{n} , $\log n$ or $20 \log n$ satisfy the $o(n)$ requirement, but in finite sample settings the actual results can vary dramatically depending on the choice. Let $\mathcal{X}_n = (X_1, \dots, X_n)$ be a random sample drawn from an unknown distribution F , and $T_n(\mathcal{X}_n, F)$ be a statistical functional of interest. Under mild conditions the m out of n bootstrap distribution $\mathcal{L}_{m,n}^*$ provides a consistent estimator of the distribution \mathcal{L}_n of $T_n(\mathcal{X}_n, F)$, provided that the bootstrap sample size m is properly chosen, refer to Götze and Račkauskas (2001) and Bickel et al. (1997). Empirical selection of m has long been an important problem, which has been discussed by, for

Empirical Bootstrapped distribution vs True empirical distribution for some m

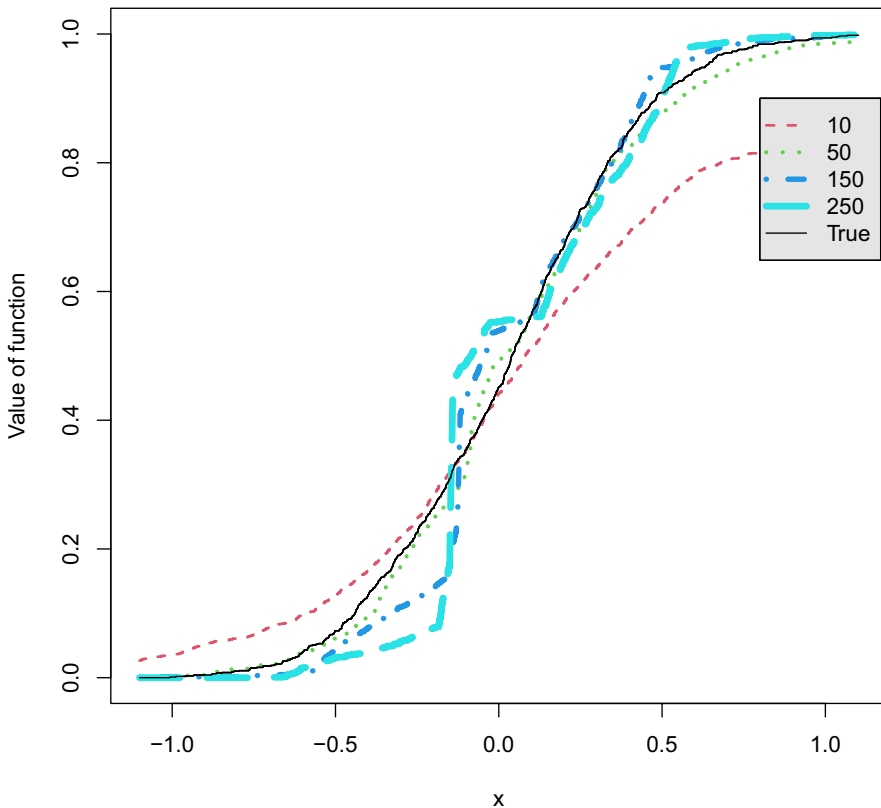


Fig. 1 Empirical distribution of $n^{1/3}(\theta_n - \theta^0)$ compared with those of $m^{1/3}(\theta_m - \theta_n)$, $m = 50$, $m = 110$, $m = 200$, $m = 250$ and $n = 250$

example, Datta and McCormick (1995), Hall et al. (1995) and Politis et al. (1999) in different contexts. The prevailing idea is to estimate a theoretically optimal sample size m , defined in a frequentist sense to be the value of m which minimises the expected value of some metric measure $d(\mathcal{L}_n, \mathcal{L}_{m,n}^*)$ between \mathcal{L}_n and $\mathcal{L}_{m,n}^*$. The problem can be solved using bootstrap samples of size m , where $m \rightarrow \infty$ and $m/n \rightarrow 0$. Bickel and Sakov (2008) proposed an adaptive rule to select a value \hat{m} and discuss its properties. The authors show, under some conditions, that $\hat{m}/n \rightarrow 1$ when the n bootstrap works, but $\hat{m} \rightarrow \infty$ and $\hat{m}/n \rightarrow 0$ when the n -bootstrap does not work. More precisely, the authors suggested the following rule for choosing m :

1. Consider a sequence of m 's of the form

$$m_j = \lfloor q^j n \rfloor, \quad \text{for } j = 0, 1, 2, \dots, \quad 0 < q < 1,$$

where $\lfloor \alpha \rfloor$ denotes the smallest integer $\geq \alpha$.

2. For each m_j , find $L_{m_j,n}^*$ (in practice this is done by Monte-Carlo).

Empirical Bootstrapped distribution vs True empirical distribution for some m

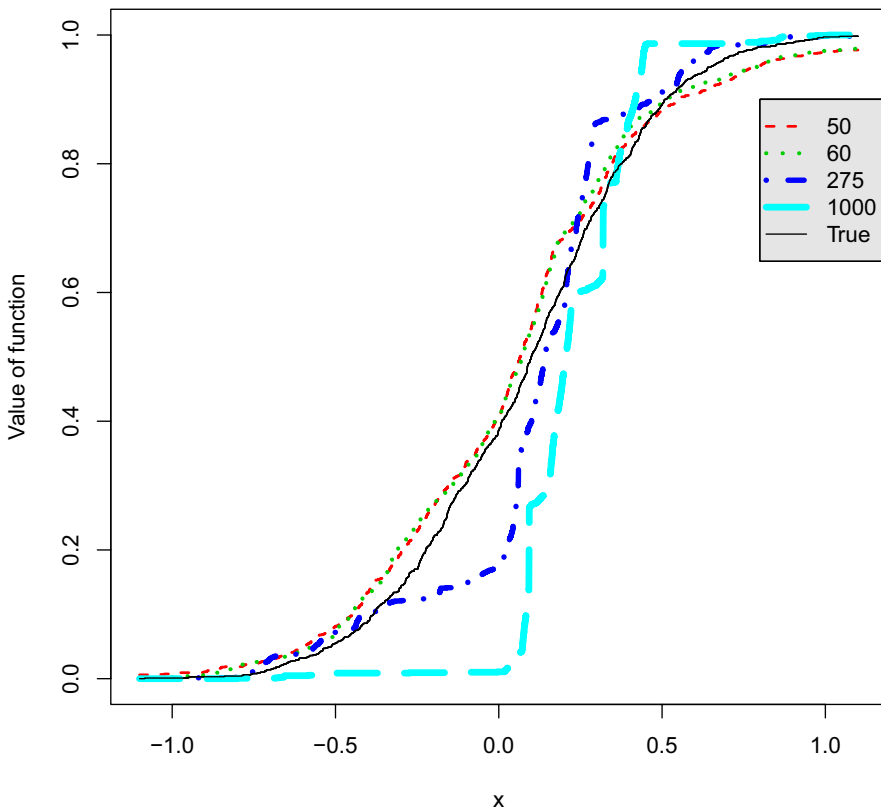


Fig. 2 Empirical distribution of $n^{1/3}(\theta_n - \theta^0)$ compared with those of $m^{1/3}(\theta_m - \theta_n)$, $m = 50$, $m = 60$, $m = 275$, $m = 1000$ and $n = 1000$

3. Let d be some metric consistent with convergence in law, and set

$$\hat{m} = \operatorname{argmin}_{m_j} d\left(L_{m_j,n}^*, L_{m_{j+1},n}^*\right).$$

If there is more than one value of m which minimizes the difference, then we pick the largest one. These results mean that the rule behaves well under both situations. Swanepoel (1986) proposed $m = (2/3)n$ to obtain the desired coverage probability of a confidence interval. Alin et al. (2017) have considered $m = n^j$ where the value j satisfies $n^j = \frac{2}{3}n$. Solving this equation for j , this expression leads to the choice

$$m = n^j \quad \text{for } j = 1 + \frac{\log\left(\frac{2}{3}\right)}{\log(n)},$$

for which we note $0 < j < 1$, so that $m = o(n)$. Götze and Račkauskas (2001) have suggested the estimation m by minimizing $d\left(\mathcal{L}_{m,n}^*, \mathcal{L}_{m/2,n}^*\right)$, yielding an optimal bootstrap sam-

Empirical Bootstrapped distribution vs True empirical distribution for some m

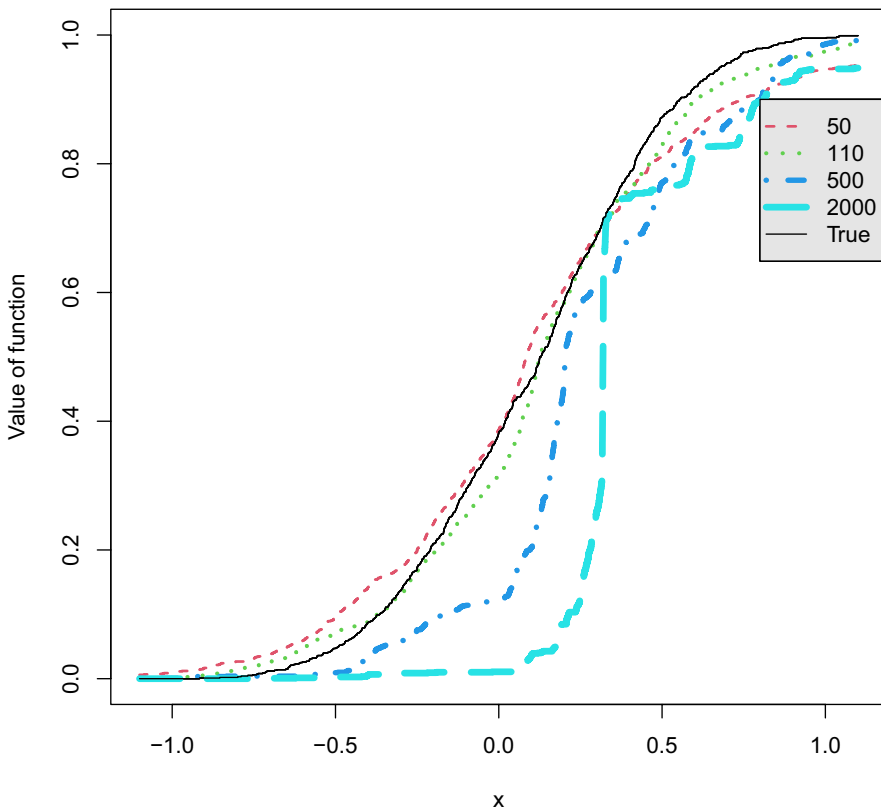


Fig. 3 Empirical distribution of $n^{1/3}(\theta_n - \theta^0)$ compared with those of $m^{1/3}(\theta_m - \theta_n)$, $m = 50$, $m = 110$, $m = 500$, $m = 2000$ and $n = 2000$

ple size in the sense of Wei et al. (2016), provided that the latter has order $o_p(n)$. Wei et al. (2016) have investigated stochastic version of the optimal bootstrap sample size, defined as the minimiser of an error measure calculated directly from the observed sample. The authors have developed procedures for calculating the stochastically optimal value of m . The performance of their methodology is illustrated in the special forms of Edgeworth-type expansions which are typically satisfied by statistics of the shrinkage type.

4 Numerical Results

We provide numerical illustrations regarding the asymptotic distribution of estimators in the classification with missing data, details are provided in Sect. 7.2. The computing program codes were implemented in R. In our simulation, we will show resampling bootstrap samples of size n fails while resampling with size m satisfying the conditions given in previous sections for the consistency of the bootstrap. Let us describe the model, define

$$X_2 = \max(\min(U + \epsilon, 1), 0),$$

where $U \sim \mathcal{U}[0, 1]$, $\epsilon \sim \mathcal{U}[-.1, .1]$ and $X_1 \sim \mathcal{U}[0, 1]$, with X_1 , ϵ and U are independent. Let

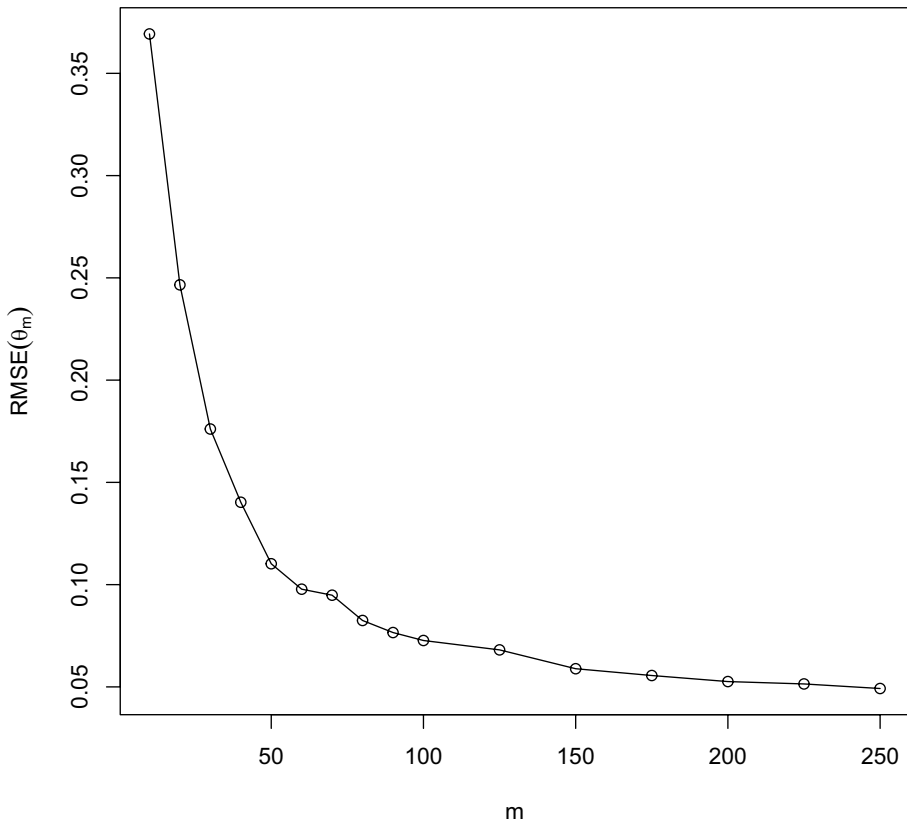


Fig. 4 The RMSE of θ_m in function of m , for $n = 250$

$$Y = \mathbb{1}\{U \geq f_\theta(X_1)\}, \tag{9}$$

where $f_\theta(x_1) = \theta x_1$, for some θ , we define

$$p(x_1) = \mathbb{P}(\Delta = 1 \mid X_1 = x_1) = \alpha_0 + (x_1 - 0.5)^2.$$

The data is composed of $Z_i = (X_{i1}, X_{i2}, Y_i \Delta_i, \Delta_i)$ $i = 1, 2, \dots, n$ from the described model. For the bandwidth, we use $h_n = \frac{c_h}{\sqrt{n}}$ ($h_m = \frac{c_h}{\sqrt{m}}$), which satisfies the requirements of regularity conditions of the asymptotic theory. In this simulation, we use the quadratic kernel defined by

$$K(u) = \frac{15}{16} (1 - u^2)^2 \mathbb{1}\{|u| \leq 1\},$$

which is a density function having support $[-1, 1]$. The results given below are based on three different value of n , we took $n = 250$, $n = 1000$ and $n = 2000$ and the true value to be $\theta^0 = 1$, we choose $c_h = 3.5$ and $\alpha_0 = 0.5$, this choice is not restrictive, we can obtain the same desired result with different value of c_h and α_0 for example $c_h = 2$ or 5 and $\alpha_0 = .25$ or $.75$. The bootstrap procedure is as follows, for each value of m we generate B

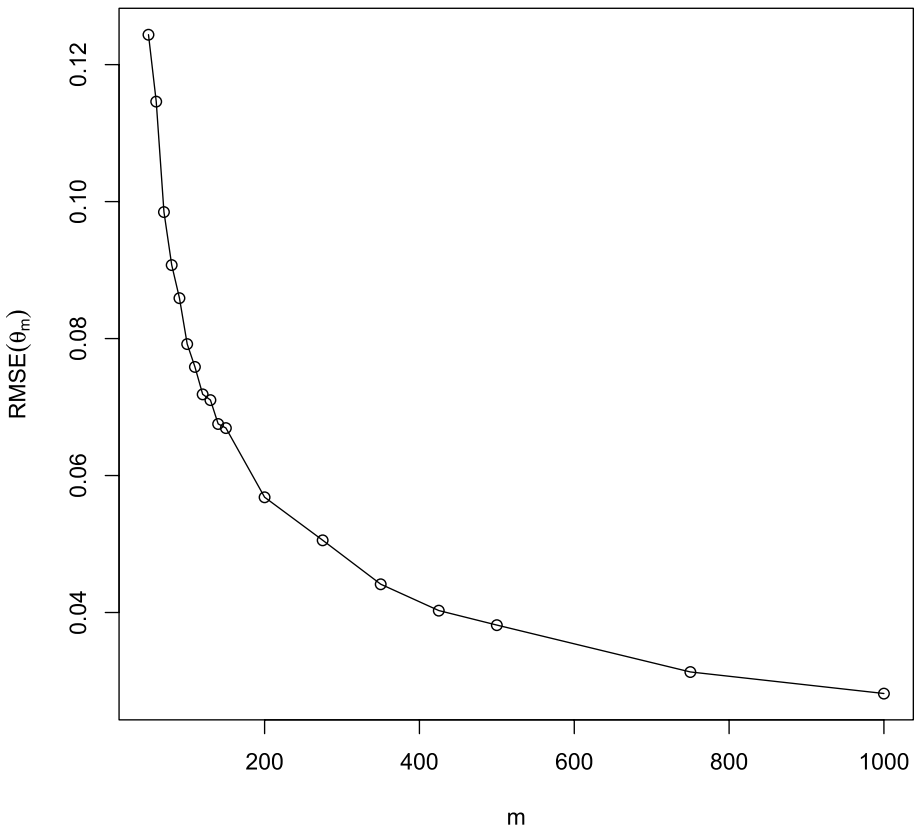


Fig. 5 The RMSE of θ_m in function of m , for $n = 1000$

independent bootstrap samples $\{Z_{ib}^* : i \leq m\}$ for $b = 1, \dots, B$, using some method of bootstrapping, and for each given value of m , we compute an estimator $\theta_m^{(b)}$ based on the b -th bootstrapped sample. Our main objective is to give a comparison between the distribution of $n^{1/3}(\theta_n - \theta^0)$ with the m out of n bootstrap distribution of $m^{1/3}(\theta_m - \theta_n)$. To achieve this goal, we have used the Kolmogorov distance between the distributions of $n^{1/3}(\theta_n - \theta^0)$ and $m^{1/3}(\theta_m - \theta_n)$ by averaging over 1000 and 1500 m out of n bootstrap sample drawn from one chosen arbitrarily random sample. Table 1 displays the results for $n = 250$, $n = 1000$ and $n = 2000$ which show that the most accurate estimates are given for the choices of $m = 50$, $m = 60$ and $m = 110$ respectively. Deviations from these choices in either direction result in deteriorating accuracy. In Figs. 1, 2 and 3, we give the empirical distribution of the true distribution and the empirical distribution of the bootstrapped one for some values of m given in Table 1, which each figure compares the estimated bootstrap empirical distribution with those of $n^{1/3}(\theta_n - \theta^0)$ for the different values of n . All these figures show that the classical bootstraps (n out n bootstrap) fail while the m out n bootstraps are consistent. Figures 4, 5 and 6 show the root mean squared error (RMSE) of the estimator θ_m for several values of m given in Table 1, for each value of n . One can see as in any other inferential context, the greater the sample size, the better.

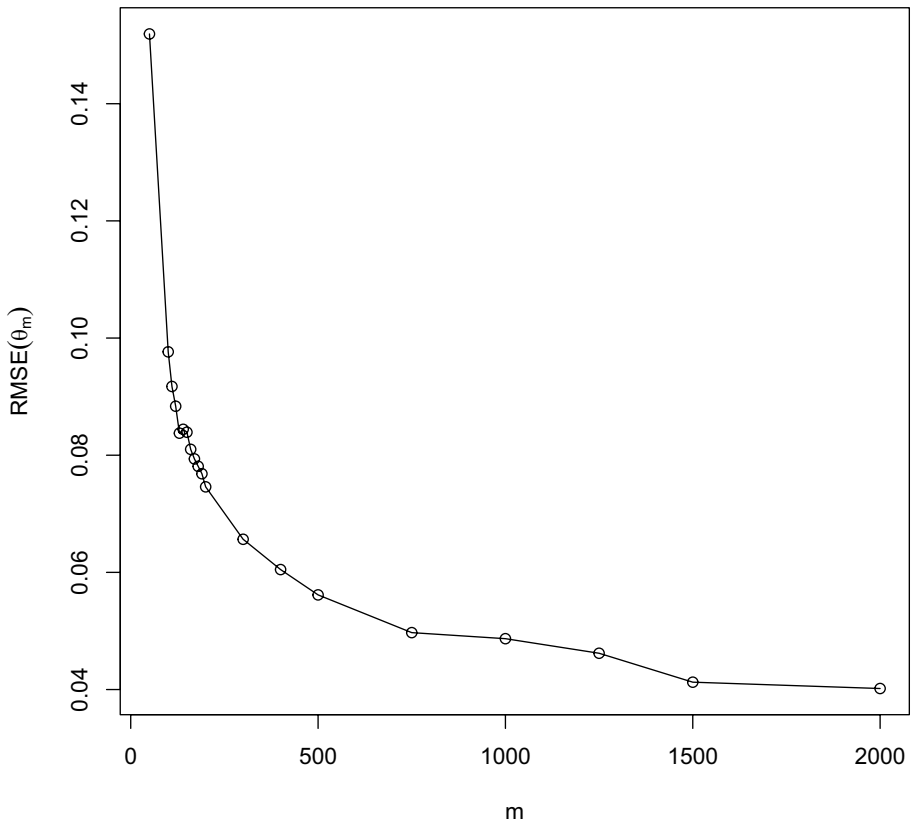


Fig. 6 The RMSE of θ_m in function of m , for $n = 2000$

5 Concluding Remarks

In the present work, we have considered the estimation of a parameter θ that maximizes a certain criterion function depending on an unknown, possibly infinite-dimensional nuisance parameter h . We have followed the common estimation procedure by maximizing the corresponding empirical criterion, in which the nuisance parameter is replaced by some nonparametric estimator. We show that the M -estimators converge weakly to maximizers of Gaussian processes in an abstract setting permitting a great flexibility for applications. We have established that the m out of n bootstrap, in this extended setting, is weakly consistent under conditions similar to those required for weak convergence of the M -estimators in the general framework of Lee (2012), when an additional difficulty comes from the nuisance parameters. The goal of this paper is therefore to extend the existing theory on the bootstrap of the M -estimators, this generalization is far from being trivial and harder to control the nuisance parameter in non-standard framework, which form a basically unsolved open problem in the literature. This requires the effective application of large sample theory techniques, which were developed for the empirical processes. Examples of applications are given to illustrate the generality and the usefulness of our results. It would be interesting to extend the results to a dependent framework, this would require further theory which is out of the scope of the present article. An important question is how to extend our findings to the setting of incomplete data (censored data, missing data, etc). This will be a subject of investigation for future work.

6 Mathematical Developments

In this section, we give the proofs of the asymptotic results of our M -estimator θ_n and its bootstrap version.

Proof of Theorem 3.2 Part (i) follows directly from Theorem 1 of Delsol and Van Keilegom (2020). For (ii), note that (A1) and (A2) imply that

$$\|\hat{\mathbb{P}}_m - \mathbb{P}\|_{\mathcal{M}_{\theta, \mathcal{H}}} = o_{\mathbb{P}_W^*}(1) \text{ a.s.} \tag{10}$$

By using the result in Lemma 3.6.16 of van der Vaart and Wellner (1996). We have; for every $\eta > 0$ there is $\delta > 0$, such that

$$\begin{aligned} & \mathbb{P}_W^* (\|\theta_m - \theta_0\| > \eta) \\ & \leq \mathbb{P}_W^* (\mathbb{P}\mathbf{m}_{\theta_0, h^0} - \mathbb{P}\mathbf{m}_{\theta_m, h^0} > \delta) \\ & \leq \mathbb{P}_W^* \left(2 \sup_{\theta \in \Theta} |\mathbb{P}\mathbf{m}_{\theta, \hat{h}_m} - \mathbb{P}\mathbf{m}_{\theta, h^0}| + \mathbb{P}\mathbf{m}_{\theta_0, \hat{h}_m} - \mathbb{P}\mathbf{m}_{\theta_m, \hat{h}_m} > \delta \right) \\ & \leq \mathbb{P}_W^* \left(2 \sup_{\theta \in \Theta} |\mathbb{P}\mathbf{m}_{\theta, \hat{h}_m} - \mathbb{P}\mathbf{m}_{\theta, h^0}| + 2\|\hat{\mathbb{P}}_m - \mathbb{P}\|_{\mathcal{M}_{\theta, \mathcal{H}}} > \delta - \hat{R}_n \right). \end{aligned}$$

Making use of the assumption (A3), there is $n_0 \in \mathbb{N}$, such that for every $n \geq n_0$, we obtain the existence of $\delta' > 0$, such that $\delta - \hat{R}_n \geq 4\delta'$ i.p., and the last expression is bounded by:

$$\begin{aligned} & \mathbb{P}_W^*(\|\theta_m - \theta_0\| > \eta) \\ & \leq \mathbb{P}_W^* \left(2 \sup_{\theta \in \Theta} |\mathbb{P} \mathbf{m}_{\theta, \hat{h}_m} - \mathbb{P} \mathbf{m}_{\theta, h^0}| + 2 \|\hat{\mathbb{P}}_m - \mathbb{P}\|_{\mathcal{M}_{\Theta, \mathcal{H}}} > 4\delta' \right) \\ & \leq \mathbb{P}_W^* \left(\sup_{\theta \in \Theta} |\mathbb{P} \mathbf{m}_{\theta, \hat{h}_m} - \mathbb{P} \mathbf{m}_{\theta, h^0}| > \delta' \right) + \mathbb{P}_W^* \left(\|\hat{\mathbb{P}}_m - \mathbb{P}\|_{\mathcal{M}_{\Theta, \mathcal{H}}} > \delta' \right). \end{aligned}$$

By using the assumptions **(AB1)**, **(A3)**, **(AB3)** in combination with **(10)**, we obtain the desired result. □

Proof of Theorem 3.5 Firstly note that, we will give the proof of this theorem for the particular choice of function

$$\psi_1(x) = Id(x) = x \text{ and } \psi_2(x) = \frac{x^\nu}{\ell(1/x)} \text{ for every } x \neq 0.$$

It worth noticing that this condition is in agreement with those used in Lee (2012) in the parametric setting. Let β_n be the $o_{\mathbb{P}^*}(1)$ in assumption **(B3)** and we define the sets

$$S_{j,n} = \{ \theta \in \Theta : 2^{j-1} < r_n \|\theta - \theta_0\| \leq 2^j \},$$

we observe

$$\Theta \setminus \theta_0 = \bigcup_{j=1}^{\infty} S_{j,n}.$$

Our objective is to show that; for any $\epsilon > 0$, there exists $\tau_\epsilon > 0$ such that

$$\mathbb{P}^*(r_n \|\theta_n - \theta_0\| > \tau_\epsilon) < \epsilon, \tag{11}$$

for any n sufficiently large. In the sequel, we work with arbitrary fixed $\epsilon > 0$. For any $\delta, \delta_1, M, K, K' > 0$, by using the condition **(A5)**, we readily obtain

$$\begin{aligned} & \mathbb{P}^*(r_n \|\theta_n - \theta_0\| > 2^M) \\ & \leq \sum_{M \leq j, 2^j \leq \delta r_n} \mathbb{P}^* \left(\sup_{\theta \in S_{j,n}} [\mathbb{P}_n \mathbf{m}_{\theta, \hat{h}} - \mathbb{P}_n \mathbf{m}_{\theta_0, \hat{h}}] \geq -K r_n^{-\nu} \ell(r_n)^{-1}, A_n \right) \\ & \quad + \mathbb{P}^*(2\|\theta_n - \theta_0\| \geq \delta) + \mathbb{P}^*(r_n^\nu \ell(r_n) |R_n| > K) + \mathbb{P}^*(r_n^{\nu-1} \ell(r_n) |W_n| > K') \\ & \quad + \mathbb{P}^* \left(|\beta_n| > \frac{C}{2} \right) + \mathbb{P}^* \left(d_{\mathcal{H}}(\hat{h}, h^0) > \frac{\delta_1}{v_n} \right), \end{aligned}$$

where

$$A_n = \left\{ r_n^{\nu-1} \ell(r_n) |W_n| \leq K', |\beta_n| \leq \frac{C}{2}, d_{\mathcal{H}}(\hat{h}, h^0) \leq \frac{\delta_1}{v_n} \right\}.$$

Indeed, we can write

$$\begin{aligned}
 & \mathbb{P}^* (r_n \|\theta_n - \theta_0\| > 2^M, 2\|\theta_n - \theta_0\| < \delta, r_n^\nu \ell(r_n) |R_n| \leq K, A_n) \\
 & \leq \sum_{j \geq M, 2^j \leq \delta r_n} \mathbb{P}^* (\theta_n \in S_{j,n}, r_n^\nu \ell(r_n) |R_n| \leq K, A_n) \\
 & \leq \sum_{j \geq M, 2^j \leq \delta r_n} \mathbb{P}^* \left(\sup_{\theta \in S_{j,n}} \left[\mathbb{P}_n \mathbf{m}_{\theta, \hat{h}} - \mathbb{P}_n \mathbf{m}_{\theta_0, \hat{h}} \right] \geq -R_n, r_n^\nu \ell(r_n) |R_n| \leq K, A_n \right) \\
 & \leq \sum_{j \geq M, 2^j \leq \delta r_n} \mathbb{P}^* \left(\sup_{\theta \in S_{j,n}} \left[\mathbb{P}_n \mathbf{m}_{\theta, \hat{h}} - \mathbb{P}_n \mathbf{m}_{\theta_0, \hat{h}} \right] \geq -K r_n^{-\nu} \ell(r_n)^{-1}, A_n \right).
 \end{aligned}$$

Condition (C) implies, for all $\delta > 0$, that there exists n_ϵ , such that, for $n > n_\epsilon$, we have

$$\mathbb{P}^* (2\|\theta_n - \theta_0\| \geq \delta) < \frac{\epsilon}{6}.$$

By the definitions of R_n , W_n and under condition (B1), there exist $\delta_1, K_\epsilon, K'_\epsilon$ and $K_{2,\epsilon}$ such that we have

$$\begin{aligned}
 & \mathbb{P}^* (r_n^\nu \ell(r_n) |R_n| > K_\epsilon) < \frac{\epsilon}{6}, \quad \mathbb{P}^* (r_n^{\nu-1} \ell(r_n) |W_n| > K'_\epsilon), \\
 & \mathbb{P}^* (|\beta_n| > \frac{C}{2}) < \frac{\epsilon}{6}, \quad \mathbb{P}^* (d_{\mathcal{H}}(\hat{h}, h^0) > \frac{\delta_1}{\nu_n}) < \frac{\epsilon}{6}.
 \end{aligned} \tag{12}$$

For n larger than some n_1 . We fix $\delta < \delta_0$ and suppose $n \geq \max(n_0, n_1, n_\epsilon)$, for $2^j \leq \delta r_n$, we have the assumptions (B2) and (B3) are fulfilled on all $S_{j,n}$. For each fixed j such that $2^j \leq \delta r_n$, under assumption (B3), for all $\theta \in S_{j,n}$, we then have

$$\begin{aligned}
 & \mathbb{P}_n \mathbf{m}_{\theta, \hat{h}} - \mathbb{P}_n \mathbf{m}_{\theta_0, \hat{h}} \\
 & \leq \mathbb{P} \mathbf{m}_{\theta, \hat{h}} - \mathbb{P} \mathbf{m}_{\theta_0, \hat{h}} + \sup_{\|\theta - \theta_0\| \leq \frac{2^j}{r_n}} \left| \mathbb{P}_n \mathbf{m}_{\theta, \hat{h}} - \mathbb{P}_n \mathbf{m}_{\theta_0, \hat{h}} - \mathbb{P} \mathbf{m}_{\theta, \hat{h}} + \mathbb{P} \mathbf{m}_{\theta_0, \hat{h}} \right| \\
 & \leq |W_n| \frac{2^j}{r_n} - (C - \beta_n) \frac{2^{\nu(j-1)}}{r_n^\nu \ell(2^{-(j-1)} r_n)} + \sup_{\|\theta - \theta_0\| \leq \frac{2^j}{r_n}} \left| \mathbb{P}_n \mathbf{m}_{\theta, \hat{h}} - \mathbb{P}_n \mathbf{m}_{\theta_0, \hat{h}} - \mathbb{P} \mathbf{m}_{\theta, \hat{h}} + \mathbb{P} \mathbf{m}_{\theta_0, \hat{h}} \right| \\
 & \leq |W_n| \frac{2^j}{r_n} - (C - \beta_n) \frac{1}{2^{-j\nu} r_n^\nu \ell(2^{-j} r_n)} + n^{-1/2} \sup_{\|\theta - \theta_0\| \leq \frac{2^j}{r_n}} \left| \mathbb{G}_n \tilde{\mathbf{m}}_{\theta - \theta_0, \hat{h}} \right|.
 \end{aligned}$$

Consequently, we obtain the following inequalities;

$$\begin{aligned}
 & \mathbb{P}^* \left(\sup_{\theta \in S_{j,n}} \left[\mathbb{P}_n \mathbf{m}_{\theta, \hat{h}} - \mathbb{P}_n \mathbf{m}_{\theta_0, \hat{h}} \right] \geq -K_\epsilon r_n^{-\nu} \ell(r_n)^{-1}, A_n \right) \\
 & \leq \mathbb{P}^* \left(n^{-1/2} \sup_{\|\theta - \theta_0\| \leq \frac{2^j}{r_n}, d_{\mathcal{H}}(h, h^0) \leq \frac{\delta_1}{\nu_n}} \left| \mathbb{G}_n \tilde{\mathbf{m}}_{\theta - \theta_0, \hat{h}} \right| \geq \frac{C}{2} \frac{2^{j\nu}}{r_n^\nu \ell(2^{-j} r_n)} - K'_\epsilon \frac{2^j}{r_n^\nu \ell(r_n)} - K_\epsilon \frac{1}{r_n^\nu \ell(r_n)} \right) \\
 & \leq \mathbb{P}^* \left(n^{-1/2} \sup_{\|\theta - \theta_0\| \leq \frac{2^j}{r_n}, d_{\mathcal{H}}(h, h^0) \leq \frac{\delta_1}{\nu_n}} \left| \mathbb{G}_n \tilde{\mathbf{m}}_{\theta - \theta_0, \hat{h}} \right| \geq \frac{2^{j\nu}}{r_n^\nu \ell(2^{-j} r_n)} \left(\frac{C}{2} - K'_\epsilon \frac{\ell(2^{-j} r_n)}{2^{j(\nu-1)} \ell(r_n)} - K_\epsilon \frac{2^{-j\nu} \ell(2^{-j} r_n)}{\ell(r_n)} \right) \right).
 \end{aligned}$$

For any $\lambda > 0$, we can find a non-decreasing function ξ such that

$$x^\lambda \ell(x) \sim \xi(x) \text{ as } x \rightarrow \infty.$$

It follows that $\frac{2^{-j\lambda} \ell(2^{-j} r_n)}{\ell(r_n)}$ is uniformly bounded for $M \leq j \leq \log_2 \delta r_n$ and for all n . Making use of the condition **(B2)** in combination with the Chebyshev’s inequality and the fact that $\varphi(c\delta) \leq c^\alpha \delta$ for all $c \geq 1$, there exists a positive constant C' and for any $\lambda > 0$, we have

$$\begin{aligned} & \mathbb{P}^* \left(\sup_{\theta \in S_{j,n}} \left[\mathbb{P}_n \mathbf{m}_{\theta, \hat{h}} - \mathbb{P}_n \mathbf{m}_{\theta_0, \hat{h}} \right] \geq -K_\epsilon r_n^{-\nu} \ell(r_n)^{-1}, A_n \right) \\ & \leq C' 2^{-j\nu} r_n^\nu \ell(2^{-j} r_n) n^{-1/2} \varphi \left(\frac{2^j}{r_n} \right) \\ & \leq C' 2^{-j(\nu-\lambda)} r_n^\nu \ell(r_n) n^{-1/2} \varphi \left(\frac{2^j}{r_n} \right) \\ & \leq C' 2^{-j(\nu-\alpha)} r_n^\nu \ell(r_n) n^{-1/2} \varphi \left(\frac{1}{r_n} \right), \end{aligned}$$

where $\nu' = \nu - \lambda > \alpha$. By choosing small value of λ and by using the properties of the function $\varphi(\cdot)$, we infer that

$$\begin{aligned} & \sum_{M \leq j, 2^j \leq \delta r_n} \mathbb{P}^* \left(\sup_{\theta \in S_{j,n}} \left[\mathbb{P}_n \mathbf{m}_{\theta, \hat{h}} - \mathbb{P}_n \mathbf{m}_{\theta_0, \hat{h}} \right] \geq -K r_n^{-\nu} \ell(r_n)^{-1}, A_n \right) \\ & \leq \sum_{M \leq j} 2^{-j(\nu'-\alpha)}, \end{aligned}$$

the last expression tends to 0 as $M \rightarrow \infty$, so we obtain the result (i) of our theorem for sufficiently large value of M and n .

For (ii) we have :

$$\begin{aligned} & \mathbb{P}_W^* (r_m \|\theta_m - \theta_0\| > 2^M) \\ & \leq \sum_{M \leq j, 2^j \leq \delta r_m} \mathbb{P}_W^* \left(\sup_{\theta \in S_{j,m}} \left[\hat{\mathbb{P}}_m \mathbf{m}_{\theta, \hat{h}_m} - \hat{\mathbb{P}}_m \mathbf{m}_{\theta_0, \hat{h}_m} \right] \geq -K r_m^{-\nu} \ell(r_m)^{-1}, A_m \right) \\ & \quad + \mathbb{P}_W^* (2\|\theta_m - \theta_0\| \geq \delta) + \mathbb{P}_W^* (r_m^\nu \ell(r_m) |\hat{R}_n| > K) + \mathbb{P}_W^* (r_m^{\nu-1} \ell(r_m) |W_m| > K') \\ & \quad + \mathbb{P}_W^* (|\beta_n| > \frac{C}{2}) + \mathbb{P}_W^* \left(d_{\mathcal{H}}(\hat{h}_m, h^0) > \frac{\delta_1}{v_m} \right). \end{aligned} \tag{13}$$

We obtain from assumption **(BB3)**, for each fixed j such that $2^j < \delta r_m$ and for all $\theta \in S_{m,j}$

$$\begin{aligned} & \widehat{\mathbb{P}}_m \widetilde{\mathbf{m}}_{\theta-\theta_0, \widehat{h}_m} \\ & \leq \mathbb{P} \widetilde{\mathbf{m}}_{\theta-\theta_0, h} + \sup_{\|\theta-\theta_0\| \leq \frac{2^j}{r_m}} \left| \widehat{\mathbb{P}}_m \widetilde{\mathbf{m}}_{\theta-\theta_0, \widehat{h}_m} - \mathbb{P}_n \widetilde{\mathbf{m}}_{\theta-\theta_0, h} + \mathbb{P}_n \widetilde{\mathbf{m}}_{\theta-\theta_0, h} - \mathbb{P} \widetilde{\mathbf{m}}_{\theta-\theta_0, h} \right| \\ & \leq |W_m| \frac{2^j}{r_m} - (C - \beta_n) \frac{1}{2^{-j\nu} r_m^\nu \ell(2^{-j} r_m)} + m^{-1/2} \sup_{\|\theta-\theta_0\| \leq \frac{2^j}{r_m}} \left| \widehat{\mathbb{G}}_m \widetilde{\mathbf{m}}_{\theta-\theta_0, \widehat{h}_m} \right| \\ & \quad + n^{-1/2} \sup_{\|\theta-\theta_0\| \leq \frac{2^j}{r_m}} \left| \mathbb{G}_n \widetilde{\mathbf{m}}_{\theta-\theta_0, \widehat{h}_m} \right|. \end{aligned}$$

This gives us, by using Chebyshev’s inequality, for some $C' > 0$

$$\begin{aligned} & \mathbb{P}_W^* \left(\sup_{\theta \in S_{j,n}} [\widehat{\mathbb{P}}_m \widetilde{\mathbf{m}}_{\theta-\theta_0, \widehat{h}_m}] \geq -K r_m^{-\nu} \ell(r_m)^{-1} \right) \\ & \leq \mathbb{P}_W^* \left(n^{-1/2} \sup_{\|\theta-\theta_0\| \leq \frac{2^j}{r_m}, d_{\mathcal{H}}(h, h^0) \leq \frac{\delta_1}{v_m}} \left| \mathbb{G}_n \widetilde{\mathbf{m}}_{\theta-\theta_0, h} \right| + m^{-1/2} \sup_{\|\theta-\theta_0\| \leq \frac{2^j}{r_m}, d_{\mathcal{H}}(h, h^0) \leq \frac{\delta_1}{v_m}} \left| \widehat{\mathbb{G}}_m \widetilde{\mathbf{m}}_{\theta-\theta_0, h} \right| \right. \\ & \quad \left. \geq \frac{2^{j\nu}}{r_m^\nu \ell(2^{-j} r_m)} \left(\frac{C}{2} - K'_\epsilon \frac{\ell(2^{-j} r_m)}{2^{j(\nu-1)} \ell(r_m)} - K_\epsilon \frac{2^{-j\nu} \ell(2^{-j} r_m)}{\ell(r_m)} \right) \right) \\ & \leq C' 2^{-j\nu} r_m^\nu \ell(r_m) m^{-1/2} \left\{ \mathbb{P}_W^* \left\| \widehat{\mathbb{G}}_m \right\|_{\mathcal{M}_{2^j/r_m, \delta_1/v_m}} + m^{1/2} n^{-1/2} \left\| \mathbb{G}_n \right\|_{\mathcal{M}_{2^j/r_m, \delta_1/v_m}} \right\}. \end{aligned}$$

From assumptions **(B2)** and **(BB2)** the outer expectation of the first term in right of (13) is bounded by

$$\begin{aligned} & \sum_{M \leq j, 2^j \leq \delta r_m} \mathbb{P} \mathbb{P}_W^* \left(\sup_{\theta \in S_{j,n}} [\widehat{\mathbb{P}}_m \mathbf{m}_{\theta, \widehat{h}_m} - \widehat{\mathbb{P}}_m \mathbf{m}_{\theta_0, \widehat{h}_m}] \geq -K r_m^{-\nu} \ell(r_m)^{-1}, A_m \right) \\ & \leq \sum_{M \leq j, 2^j \leq \delta r_m} C' 2^{-j\nu} r_m^\nu \ell(r_m) m^{-1/2} \left\{ \mathbb{P} \mathbb{P}_W^* \left\| \widehat{\mathbb{G}}_m \right\|_{\mathcal{M}_{2^j/r_m, \delta_1/v_m}} + m^{1/2} n^{-1/2} \mathbb{P} \left\| \mathbb{G}_n \right\|_{\mathcal{M}_{2^j/r_m, \delta_1/v_m}} \right\} \\ & \leq \sum_{M \leq j, 2^j \leq \delta r_m} C' 2^{-j\nu} r_m^\nu \ell(r_m) m^{-1/2} \varphi \left(\frac{2^j}{r_m} \right) \\ & \quad + \sum_{M \leq j, 2^j \leq \delta r_m} C' m^{1/2} n^{-1/2} 2^{-j\nu} r_m^\nu \ell(r_m) m^{-1/2} \varphi \left(\frac{2^j}{r_m} \right) \\ & \leq C' \sum_{M \leq j} 2^{-j(\nu-\alpha)} + C' m^{1/2} n^{-1/2} \sum_{M \leq j} 2^{-j(\nu-\alpha)}, \end{aligned}$$

with assumption **(AB1)** in mind the last two terms converge to 0 as $M, n \rightarrow \infty$, the outer expectation of the others terms in (13) are $o_{\mathbb{P}_W^*}(1)$ i.p., by Lemma 3 of Cheng and Huang (2010), which completes the proof of Theorem 3.5. □

Proof of Theorem 3.8 The proof of the first part (i) of Theorem 3.8 is given in Lemmas 1, 2 and 3 of Delsol and Van Keilegom (2020), where in our setting we use bounded closed subsets in the place of compact subsets. We note by their Lemma 2, we obtain the existence of $\xi_{1,n}, \xi_{2,n}, \xi_{3,n}$ such that

$$\sup_{\gamma \in \mathcal{K}} |\xi_{l,n}| = o_{\mathbb{P}}(1), \text{ for } l = 1, 2, 3,$$

and the following decomposition

$$r_n^\nu \ell(r_n) \mathbb{P}_n \tilde{\mathbf{m}}_{r_n, \hat{h}}(1 + \xi_{1,n}) = \left[r_n^{\nu-1} \ell(r_n) W_n + r_n^\nu \ell(r_n) \mathbb{P}_n \tilde{\mathbf{m}}_{r_n, h^0} \right] (1 + \xi_{2,n}) + \xi_{3,n}.$$

By their Lemma 3, the properties of the function $\gamma \mapsto W_n(\gamma)$ and the assumptions of Theorem 3.8; we obtain the weak convergence of the process

$$\gamma \mapsto r_n^{\nu-1} \ell(r_n) W_n(\gamma) + r_n^\nu \ell(r_n) \mathbb{P}_n \tilde{\mathbf{m}}_{r_n, h^0}.$$

Briefly, we have the following decomposition;

$$T_n(\gamma) = r_n^{\nu-1} \ell(r_n) W_n(\gamma) + r_n^\nu \ell(r_n) \mathbb{P}_n \tilde{\mathbf{m}}_{r_n, h^0} = T_{1,n}(\gamma) + T_{2,n}(\gamma),$$

where

$$T_{1,n}(\gamma) = \mathbb{M}_n(\gamma, h^0)$$

and

$$T_{2,n} = r_n^\nu \ell(r_n) \mathbb{P}_n \tilde{\mathbf{m}}_{r_n, h^0} + r_n^{\nu-1} \ell(r_n) W_n(\gamma).$$

The process $\gamma \mapsto T_{1,n}(\gamma)$ does not depend on the estimation of nuisance parameter, so it can be studied in a similar way as in the parametric model, by Theorem 2.11.1 of van der Vaart and Wellner (1996) and the use of assumptions (C4), (C5), (C9) and (C10), we obtain its uniformly asymptotic equicontinuity. For the process $\gamma \mapsto T_{2,n}(\gamma)$, we can show that it is asymptotically uniformly equicontinuous by the same method given in the proof of their Lemma 3. By Theorem 1.5.7 and 1.5.4 of van der Vaart and Wellner (1996), we obtain the asymptotic tightness and the weak convergence of T_n to $\Lambda + \mathbb{G}$ in $\ell(\mathcal{K})$ and using Addendum 1.5.8 in the same reference; the almost all paths of the limiting process on \mathcal{K} are uniformly continuous with respect to $\|\cdot\|$. Finally by Slutsky’s theorem we obtain the desired result. \square

For part (ii) we are in the situation to show the weak convergence of the bootstrapped process, which follows directly from Slutsky’s theorem and Lemmas 6.1 and 6.2 given below.

Lemma 6.1 *Let $\mathcal{K} = \{\gamma \in \mathbf{E} : \|\gamma\| \leq K\}$. Then under assumptions of part (ii) of Theorem 3.8, for all $\gamma \in \mathcal{K}$, there exist $z_{0,m}, z_{1,m}, z_{2,m}$, such that*

$$\sup_{\gamma \in \mathcal{K}} |z_{j,m}| = o_{\mathbb{P}_W^*}(1), \text{ i.p. , } j = 0, 1, 2,$$

and

$$\begin{aligned} & r_m^\nu \ell(r_m) \hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, \hat{h}_m}(1 + z_{0,m}) \\ &= \left[r_m^\nu \ell(r_m) \hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, h^0} + r_m^{\nu-1} \ell(r_m) W_m(\gamma) \right] (1 + z_{1,m}) + z_{2,m}. \end{aligned}$$

Proof of Lemma 6.1 We need to introduce the following notation

$$\alpha_{0,n}(\gamma) = \frac{\mathbb{P}_n \tilde{\mathbf{m}}_{r_m, \hat{h}_n} - \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, \hat{h}_n} - \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, h^0} + \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, h^0}}{r_n^{-\nu} \ell^{-1}(r_n) + \left| \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, \hat{h}_n} \right| + \left| \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, h^0} \right| + \left| \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, \hat{h}_n} \right| + \left| \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, h^0} \right|},$$

$$\alpha_{0,m}(\gamma) = \frac{\hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, \hat{h}_n} - \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, \hat{h}_n} - \hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, h^0} + \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, h^0}}{r_m^{-\nu} \ell^{-1}(r_m) + \left| \hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, \hat{h}_n} \right| + \left| \hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, h^0} \right| + \left| \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, \hat{h}_n} \right| + \left| \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, h^0} \right|},$$

$$s_{n,h}(\gamma) = \text{sign} \left[\mathbb{P}_n \tilde{\mathbf{m}}_{r_n, h} \right],$$

$$s_h(\gamma) = \text{sign} \left[\mathbb{P} \tilde{\mathbf{m}}_{r_n, h} \right],$$

$$s_{m,h}(\gamma) = \text{sign} \left[\hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, h} \right].$$

The set \mathcal{K} is bounded and θ_0 belongs to the interior of Θ , there exist $m_{\mathcal{K}}$ such that for all $m \geq m_{\mathcal{K}}$ and for all $\gamma \in \mathcal{K}$, the quantity $\theta_0 + \frac{\gamma}{r_m}$ is in Θ . Then for all $\gamma \in \mathcal{K}$ we have;

$$\begin{aligned} \hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, \hat{h}_n} &= \hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, h^0} + \mathbb{P} \tilde{\mathbf{m}}_{r_m, \hat{h}_n} - \mathbb{P} \tilde{\mathbf{m}}_{r_n, h^0} \\ &+ \alpha_{0,m}(\gamma) \left(r_m^{-\nu} \ell^{-1}(r_m) + \left| \hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, \hat{h}_n} \right| \right. \\ &+ \left. \left| \hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, h^0} \right| + \left| \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, \hat{h}_n} \right| + \left| \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, h^0} \right| \right) \\ &+ \alpha_{0,n}(\gamma) \left(r_n^{-\nu} \ell^{-1}(r_n) + \left| \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, \hat{h}} \right| + \left| \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, h^0} \right| + \left| \mathbb{P} \tilde{\mathbf{m}}_{r_m, \hat{h}} \right| + \left| \mathbb{P} \tilde{\mathbf{m}}_{r_m, h^0} \right| \right). \end{aligned}$$

This can be rewritten as follows

$$\begin{aligned} &r_m^\nu \ell(r_m) \hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, \hat{h}_n} \left(1 - \alpha_{0,m}(\gamma) s_{m, \hat{h}_n}(\gamma) \right) \\ &= r_m^\nu \ell(r_m) \hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, h^0} \left(1 + \alpha_{0,m}(\gamma) s_{m, h^0}(\gamma) \right) \\ &+ r_m^\nu \ell(r_m) \mathbb{P} \tilde{\mathbf{m}}_{r_m, \hat{h}_n} \left(1 + \alpha_{0,n}(\gamma) s_{\hat{h}_n}(\gamma) \right) \\ &- r_m^\nu \ell(r_m) \mathbb{P} \tilde{\mathbf{m}}_{r_m, h^0} \left(1 - \alpha_{0,n}(\gamma) s_{h^0}(\gamma) \right) + z'_{2,m}, \end{aligned} \tag{14}$$

where;

$$\begin{aligned} z'_{2,m} &= \alpha_{0,m}(\gamma) + r_m^\nu \ell(r_m) (\alpha_{0,m}(\gamma) + \alpha_{0,n}(\gamma)) \left| \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, \hat{h}_n} \right| \\ &+ r_m^\nu \ell(r_m) (\alpha_{0,m}(\gamma) + \alpha_{0,n}(\gamma)) \left| \mathbb{P}_n \tilde{\mathbf{m}}_{r_m, h^0} \right| + \frac{r_m^\nu \ell(r_m)}{r_n^\nu \ell(r_n)} \alpha_{0,n}(\gamma). \end{aligned}$$

We get from the assumption **(CB1)** and **(CB3)** that

$$\begin{aligned}
 r_m^\nu \ell(r_m) \left[\mathbb{P} \tilde{\mathbf{m}}_{r_m, \hat{h}_n} - \mathbb{P} \tilde{\mathbf{m}}_{r_m, h^0} \right] &= r_m^{\nu-1} \ell(r_m) W_m(\gamma) + r_m^{\nu-2} \ell(r_m) \beta_n \|\gamma\|^2 \\
 &\quad + r_m^{\nu-2} \ell(r_m) o(\|\gamma\|^2) \\
 &:= r_m^{\nu-1} \ell(r_m) W_m(\gamma) + \alpha_{1,n}(\gamma).
 \end{aligned}
 \tag{15}$$

By combining (14) and (15), we infer that

$$\begin{aligned}
 &r_m^\nu \ell(r_m) \widehat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, \hat{h}_n} (1 + z_{0,m}) \\
 &= \left[r_m^\nu \ell(r_m) \widehat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, h^0} + r_m^{\nu-1} \ell(r_m) W_m(\gamma) \right] (1 + z_{1,m}) + z_{2,m},
 \end{aligned}$$

where

$$\begin{aligned}
 z_{0,m}(\gamma) &= -\alpha_{0,m}(\gamma) s_{m, \hat{h}_n}(\gamma) \\
 z_{1,m}(\gamma) &= \alpha_{0,m}(\gamma) s_{m, h^0}(\gamma) \\
 z_{2,m}(\gamma) &= z'_{2,m}(\gamma) + z''_{2,m}(\gamma),
 \end{aligned}$$

and

$$\begin{aligned}
 z''_{2,m}(\gamma) &= \alpha_{0,n}(\gamma) \left[1 + (V(\gamma, \gamma) + r_m^{\nu-2} \ell(r_m) o(\|\gamma\|^2)) (s_{\hat{h}_n} + s_{h_0})(\gamma) \right. \\
 &\quad \left. + (r_m^{\nu-1} \ell(r_m) W_m(\gamma) + \alpha_{1,n}(\gamma)) (s_{\hat{h}_n} - s_{n, h_0})(\gamma) \right] \\
 &\quad + \alpha_{1,n}(\gamma) (1 + z_{1,m}(\gamma)).
 \end{aligned}$$

It is easily to show that

$$\sup_{\gamma \in \mathcal{K}} |z_{j,m}| = o_{\mathbb{P}_w^*}(1) \text{ i.p., for } j = 0, 1, 2,$$

by using assumptions (A2), (AB1) (C3), (CB2), (CB3) and Lemma 3 of Cheng and Huang (2010). □

Lemma 6.2 *Under the assumptions of Lemma 6.1, the process*

$$\gamma \mapsto r_m^\nu \ell(r_m) \widehat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, \hat{h}_n} + r_m^{\nu-1} \ell(r_m) W_m(\gamma)$$

converges weakly conditionally in probability to the process

$$\gamma \mapsto \Lambda(\gamma) + \mathbb{G}(\gamma) \text{ in } \ell^\infty(\mathcal{K}).$$

Proof of Lemma 6.2 Making use of the assumption (CB4), we need only to show the equi-continuity of the process

$$T_m : \gamma \mapsto r_m^\nu \ell(r_m) \widehat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, \hat{h}_n} + r_m^{\nu-1} \ell(r_m) W_m(\gamma).$$

One can see that the process T_m can be decomposed into the sum three processes in the following way

$$T_m = \sum_{i=1}^3 T_{i,m},$$

where

$$\begin{aligned} T_{1,m} &: \gamma \mapsto r_m^\nu \ell(r_m) \left((\hat{\mathbb{P}}_m - \mathbb{P}_n) \tilde{\mathbf{m}}_{r_m, h^0} \right), \\ T_{2,m} &: \gamma \mapsto r_m^\nu \ell(r_m) \left((\mathbb{P}_n - \mathbb{P}) \tilde{\mathbf{m}}_{r_m, h^0} \right), \\ T_{3,m} &: \gamma \mapsto r_m^\nu \ell(r_m) \mathbb{P} \tilde{\mathbf{m}}_{r_m, h^0} + r_m^{\nu-1} \ell(r_m) W_m(\gamma). \end{aligned}$$

We shall study separately the properties of each process. Firstly, we note that by assumption **(C6)**, **(CB3)** and **(AB1)**, for sufficiently large m , we have $\theta_0 + \frac{\mathcal{K}}{r_m} \subset \Theta$ and then the processes $T_{1,m}$, $T_{2,m}$ and $T_{3,m}$ take values in $\ell^\infty(\mathcal{K})$. The process $T_{2,m}$ can be treated as in the proof of part **(i)** by reformulating it to this form

$$\begin{aligned} T_{2,m}(\gamma) &= r_m^\nu \ell(r_m) \left((\mathbb{P}_n - \mathbb{P}) \tilde{\mathbf{m}}_{r_m, h^0} \right) \\ &= \sqrt{\frac{m}{n}} \mathbb{G}_n \frac{r_m^\nu \ell(r_m)}{\sqrt{m}} \tilde{\mathbf{m}}_{r_m, h^0}, \end{aligned} \tag{16}$$

as in the proof of **(i)** apply Theorem 2.11.22 of van der Vaart and Wellner (1996) to the process

$$\gamma \mapsto \mathbb{G}_n \frac{r_m^\nu \ell(r_m)}{\sqrt{m}} \tilde{\mathbf{m}}_{r_m, h^0},$$

by assumptions **(C4)**, **(C5)**, **(C9)** and **(C10)** we get its uniform equicontinuity with respect to $\|\cdot\|$ on \mathcal{K} and by the use of assumption **(AB1)**, we obtain our main result for the process $T_{2,m}$. Then the process $T_{1,m}$ also does not depend to the estimation of the nuisance parameter, it can be treated in the same way as in part **(ii)** of Lemma 2 in Lee (2012). Briefly, we want to show that

$$\Delta_n \stackrel{\text{def}}{=} P_W^* \sup \{ |T_{1,m}(\gamma) - T_{1,m}(\gamma')| : \|\gamma - \gamma'\| \leq \delta_n, \gamma, \gamma' \in \mathcal{K} \} \rightarrow 0 \text{ i.p.}$$

Define the class

$$\mathcal{H}_n = r_m^\nu \ell(r_m) m^{-1/2} \mathcal{M}_{d/r_m}(\delta_n/r_m),$$

and let

$$\tilde{M}_n = r_m^\nu \ell(r_m) m^{-1/2} M_{d/r_m}$$

its envelope function. Making use of the condition **(B2)**, we readily infer that

$$P^* \sup \{ |f| : f \in \mathcal{H}_n \} \leq C P^* \tilde{M}_n \leq C < \infty.$$

It follows by Lemmas 2.9.1 and 3.6.6 of van der Vaart and Wellner (1996) that, for $1 \leq n_0 \leq n$;

$$\begin{aligned} \mathbb{P}^* \Delta_n \leq & C n_0 m^{-1/2} \left(\mathbb{E} \max_{1 \leq i \leq n} \tilde{N}_i \right) \mathbb{P}^* \tilde{M}_n \\ & + C n^{1/2} m^{-1/2} \int_0^\infty \left(\mathbb{P} \{ |\tilde{N}_1| > x \} \right)^{1/2} dx \max_{n_0 \leq j \leq n} \mathbb{P}^* \left\| j^{-1/2} \sum_{i=n_0}^j \varepsilon_i \delta_{\mathbf{X}_i} \right\|_{\mathcal{H}_n}, \end{aligned}$$

where $\tilde{N} = (\tilde{N}_1, \tilde{N}_2, \dots)$ and $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots)$ are independent sequences of i.i.d. symmetrized Poisson variables with parameter $m/(2n)$ and i.i.d. Rademacher variables, respectively and both being independent of $\mathbf{X}_1, \dots, \mathbf{X}_n$. By Jensen’s inequality, problem 3.6.3 of van der Vaart and Wellner (1996) and (B2), we readily get

$$\mathbb{P}^* \Delta_n \leq C n_0 m^{-1/2} \log n / \log(n/m + 1) + C \max_{n_0 \leq j \leq n} \mathbb{P}^* \left\| j^{-1/2} \sum_{i=n_0}^j \varepsilon_i \delta_{\mathbf{X}_i} \right\|_{\mathcal{H}_n}. \tag{17}$$

By taking

$$n_0 = n_{0,n} = a_m m^{1/4} \{ \log(n/m + 1) / \log n \}^{1/2} \in [1, n],$$

it follows, by condition (C4), that

$$n_0 m^{-1/2} \log n / \log(n/m + 1) \rightarrow 0 \quad \text{as } n \rightarrow \infty. \tag{18}$$

We refer to the integrand in (C9) by $\mathcal{N}(\varepsilon)$ for $\varepsilon > 0$. By using the triangular inequality, the properties of sub-Gaussianity of Rademacher processes, under (C10) and using the Cauchy-Schwarz inequality with (B2), we obtain

$$\begin{aligned} \max_{n_0 \leq j \leq n} \mathbb{P}^* \left\| j^{-1/2} \sum_{i=n_0}^j \varepsilon_i \delta_{\mathbf{X}_i} \right\|_{\mathcal{H}_n} &= \max_{n_0 \leq j \leq n} \mathbb{P}_X \mathbb{P}_\varepsilon \left\| j^{-1/2} \sum_{i=n_0}^j \varepsilon_i \delta_{\mathbf{X}_i} \right\|_{\mathcal{H}_n} \\ &\leq 2 \max_{n_0 \leq j \leq n} \mathbb{P}_X \mathbb{P}_\varepsilon \left\| j^{-1/2} \sum_{i=1}^j \varepsilon_i \delta_{\mathbf{X}_i} \right\|_{\mathcal{H}_n} \\ &\leq C \max_{n_0 \leq j \leq n} \left\{ \mathbb{P}^* \left(\int_0^{\psi_{n,j}} \mathcal{N}(\varepsilon) d\varepsilon \right)^2 \right\}^{1/2} (\mathbb{P}^* \tilde{M}_n^2)^{1/2} \\ &\leq C \max_{n_0 \leq j \leq n} \left\{ \mathbb{P}^* \left(\int_0^{\psi_{n,j}} \mathcal{N}(\varepsilon) d\varepsilon \right)^2 \right\}^{1/2}, \end{aligned} \tag{19}$$

where

$$\psi_{n,j} = \sup \left\{ \|f\|_{\mathbb{P}_{j,2}} : f \in \mathcal{H}_n \right\}.$$

Our aim is to show that

$$\psi_{n_k, j_k} = o_{\mathbb{P}^*}(1) \quad \text{as } k \rightarrow \infty, \tag{20}$$

for an arbitrary subsequence $\{n_k : k = 1, 2, \dots\}$ of $\{n\}$, and any arbitrary sequence $\{j_k\}$ such that $n_{0,n_k} \leq j_k \leq n_k$ for all $k = 1, 2, \dots$. Write $m_k^* = m_{n_k}$. Define, for any $\gamma \in \mathcal{K}$,

$$Z_{ki}(\gamma) = r_m^{v_k} \ell(r_m^{*k}) m_k^{*-1/2} j_k^{-1/2} \tilde{m}_{\gamma/r_m^{*k}}(\mathbf{X}_i).$$

As in the proof of part (ii) of Lemma 2 of Lee (2012); he showed these variables satisfy the condition of Theorem 2.11.1 of van der Vaart and Wellner (1996), which implies our result in (20) for arbitrary subsequence n_k and $j_k \in [n_{0,n_k}, n]$, by arguing as in the proof of this Theorem. It then follows by the dominated convergence theorem that the bound in (19) has limsup equal to 0 as $n \rightarrow \infty$. Substituting this and (18) into (17) to obtain the desired result. Finally, for the process $T_{3,m}$, for large value of m , we have $\theta_0 + \frac{\gamma}{r_m} \in \Theta$ by using the assumption (CB3), we get, for all $0 < \delta < \delta_1$,

$$\begin{aligned} & \sup_{\gamma, \gamma' \in \mathcal{K}, \|\gamma - \gamma'\| \leq \delta} |T_{3,m}(\gamma) - T_{3,m}(\gamma')| \\ &= \sup_{\gamma, \gamma' \in \mathcal{K}, \|\gamma - \gamma'\| \leq \delta} \left| r_m^{v-1} \ell(r_m) W_m(\gamma - \gamma') + r_m^{v-2} \ell(r_m) (V(\gamma, \gamma) - V(\gamma', \gamma')) \right. \\ & \quad \left. + r_m^v \ell(r_m) \left(o\left(\frac{\|\gamma\|^2}{r_m^2}\right) + o\left(\frac{\|\gamma'\|^2}{r_m^2}\right) \right) \right| \\ &\leq \delta^\tau \left(r_m^{v-1} \ell(r_m) \sup_{\gamma \in \mathcal{K}, \delta \leq \delta_1, \|\gamma\| \leq \delta} \left| \frac{W_m(\gamma)}{\delta^\tau} \right| + r_m^{v-2} \ell(r_m) \sup_{\gamma, \gamma' \in \mathcal{K}, \delta \leq \delta_1, \|\gamma - \gamma'\| \leq \delta} \frac{|V(\gamma, \gamma) - V(\gamma', \gamma')|}{\delta^\tau} \right) \\ & \quad + b_m \\ &:= \delta^\tau \alpha_m + b_m, \end{aligned} \tag{21}$$

where

$$b_m \leq \sup_{\gamma, \gamma' \in \mathcal{K}} r_m^v \ell(r_m) \left(o\left(\frac{\|\gamma\|^2}{r_m^2}\right) + o\left(\frac{\|\gamma'\|^2}{r_m^2}\right) \right) \rightarrow 0, \text{ as } m \rightarrow \infty,$$

and $\alpha_m = O_{\mathbb{P}_W^*}(1)$ i.p. uniformly over $\delta \leq \delta_1$. From this, we obtain, for any $\epsilon > 0$ and $\eta > 0$,

$$\begin{aligned} & \mathbb{P}_W^* \left(\sup_{\gamma, \gamma' \in \mathcal{K}, \|\gamma - \gamma'\| \leq \delta} |T_{3,m}(\gamma) - T_{3,m}(\gamma')| > \epsilon \right) \\ & \leq \mathbb{P}_W^* \left(\delta^\tau \alpha_m + b_m > \epsilon, \alpha_m \leq C, |b_m| < \frac{\epsilon}{2} \right) + \mathbb{P}_W^* (\alpha_m > C) \\ & \leq \mathbb{P}_W^* \left(\delta^\tau > \frac{\epsilon}{2C} \right) + \mathbb{P}_W^* (\alpha_m > C). \end{aligned}$$

By choosing C_η such that the last term is bounded by η for large value of m , and taking $\delta \leq \delta_1 \wedge \left(\frac{\epsilon}{2C_\eta}\right)^\frac{1}{\tau}$, which implies the main result for the process $T_{3,m}$. Finally by using the fact that

$$T_m = T_{1,m} + T_{2,m} + T_{3,m},$$

we obtain the desired result on the process T_m . □

Proof of Theorem 3.9 Making use of the result (i) in Theorem 3.8 in connection with the assumption (C8), we infer that we have almost all paths of the process $\gamma \mapsto \mathbb{G}(\gamma) + \Lambda(\gamma)$ are uniformly continuous on every $\mathcal{K} \subset \mathcal{B}$, and reaching the supremum at an unique point γ_0 . For part (i), an application of (i) in Theorem 3.8, for any closed bounded $\mathcal{K} \subset \mathcal{B}$, gives

$$\mathbb{T}_n(\gamma) = r_n^\vee \ell(r_n) \mathbb{P}_n \tilde{\mathbf{m}}_{\underline{x}, \hat{h}_n} \Rightarrow \Lambda(\gamma) + \mathbb{G}(\gamma), \text{ in } \ell^\infty(\mathcal{K}).$$

We get from the assumption **(C11)** that

$$\mathbb{T}_n(r_n(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)) \geq \sup_{\gamma \leq K} \mathbb{T}_n(\gamma) - o_{\mathbb{P}}(1).$$

Noting that γ_0 is the unique, well-separated, maximizer of $\mathbb{G}(\gamma) + \Lambda(\gamma)$, then part **(i)** follows by Theorem 3.2.2 of van der Vaart and Wellner (1996), where compact sets and uniform tightness of $r_n(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)$ are replaced respectively by closed bounded sets with similar structure as the set \mathcal{K} and

$$r_n(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) = O_{\mathbb{P}^*}(1).$$

For part **(ii)**, we infer that

$$\hat{\mathbb{T}}_m(\gamma) = r_m^\vee \ell(r_m) \hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{\underline{x}, \hat{h}_m} \Rightarrow \Lambda(\gamma) + \mathbb{G}(\gamma) \text{ i.p. in } \ell(\mathcal{K}).$$

By combining the assumption **(CB5)** with the first part of **(CB1)**, we have respectively

$$\hat{\mathbb{T}}_m(r_m(\boldsymbol{\theta}_m - \boldsymbol{\theta}_0)) \geq \sup_{\gamma \leq K} \hat{\mathbb{T}}_m(\gamma) - o_{\mathbb{P}^*}(1)$$

and

$$r_m(\boldsymbol{\theta}_m - \boldsymbol{\theta}_0) = O_{\mathbb{P}^*}(1), \text{ i.p.}$$

An application of Lemma 4(ii) of Lee (2012) gives

$$r_m(\boldsymbol{\theta}_m - \boldsymbol{\theta}_0) \Rightarrow \gamma_0, \text{ i.p.}$$

It follows from the first part of the assumptions **(C1)**, **(AB1)** and Slutsky’s theorem that

$$r_m(\boldsymbol{\theta}_m - \boldsymbol{\theta}_n) = r_m(\boldsymbol{\theta}_m - \boldsymbol{\theta}_0) - \frac{r_m}{r_n} r_n(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) \Rightarrow \gamma_0 \text{ i.p.}$$

Hence the proof of the statement **(ii)** is complete. □

7 Appendix: Applications

We present in this section some examples which can not handled with the classical theory of semiparametric estimators and their m out of n bootstrap version cannot be applied while theory of the paper can be applied. This illustrate the usefulness of our results. Delsol and Van Keilegom (2020) provided some examples of situations in which the existing theory on semiparametric estimators cannot be applied, whereas their result could be applied. It worth noticing that the aim of this section is to verify the bootstrap conditions that are different from those used for the non bootstrapped estimators checked in the last mentioned reference. Although only three examples will be given here, they stand as archetypes for a variety of models that can be investigated by the methodology of the present paper.

7.1 Single Index Model with Monotone Link Function

The single index regression models are typical examples which are given

$$Y = g(\mathbf{X}^\top \boldsymbol{\beta}) + \varepsilon \tag{22}$$

where $\mathbb{P}(\varepsilon|\mathbf{X}) = 0$, $\text{Var}(\varepsilon|\mathbf{X}) < \infty$ and we assume that the unknown function $g(\cdot)$ is monotone, we refer to Ichimura (1993) for more details. On the basis of the sample $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ coming from the model (22), we make use of the the pool-adjacent-violators algorithm to construct and estimator of the function $g(\cdot)$. This gives a non-smooth estimator $\hat{g}_\beta(\cdot)$ of $g_\beta(\mathbf{z}) = \mathbb{E}[Y|\mathbf{X}^\top \boldsymbol{\beta} = \mathbf{z}]$. Next, by using the least-squares estimation method we estimate $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = \arg \max_{\boldsymbol{\beta}} \left[-n^{-1} \sum_{i=1}^n (Y_i - \hat{g}_\beta(\mathbf{X}_i^\top \boldsymbol{\beta}))^2 \right].$$

By the fact that $\hat{g}_\beta(\cdot)$ is of non-smooth nature implies that the criterion function is not smooth in $\boldsymbol{\beta}$. This is a situation where the theory of the present paper can be applied.

7.2 Classification with Missing Data

Let $\mathbf{X}_1 = (\mathbf{X}_{11}, \mathbf{X}_{12}), \dots, \mathbf{X}_n = (\mathbf{X}_{n1}, \mathbf{X}_{n2})$ be independent and identically distributed random copies of the random vector $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, coming from two underlying populations. For $j = 0, 1$, let $\mathbf{Y}_i = j$ when the \mathbf{X}_i comes from the population j . Let us denote by \mathbf{Y} the population indicator associated with the vector \mathbf{X} . Using the information of available data, we seek to find a classification method for novel observations with unknown true population.

The classification is performed by regressing \mathbf{X}_2 on \mathbf{X}_1 making use of the parametric criterion function $f_\theta(\cdot)$, and choosing θ that maximize the following

$$\mathbb{P} \mathbb{I}_{\{\mathbf{Y}=1, \mathbf{X}_2 \geq f_\theta(\mathbf{X}_1)\}} + \mathbb{P} \mathbb{I}_{\{\mathbf{Y}=0, \mathbf{X}_2 < f_\theta(\mathbf{X}_1)\}}. \tag{23}$$

Let θ_0 denote the maximizer of (23) with respect to all $\theta \in \Theta$, here Θ is assumed to be a compact subset of \mathbb{R}^k containing an interior point θ_0 . Now assume that \mathbf{Y}_i 's are subject to some missing mechanism. Let Δ_i be a random variable (respectively Δ) equals to 1 when we observe the random variable \mathbf{Y}_i (respectively \mathbf{Y}), and 0 otherwise. Let $\mathbf{Z}_1 = (\mathbf{X}_1, \mathbf{Y}_1 \Delta_1, \Delta_1), \dots, \mathbf{Z}_n = (\mathbf{X}_n, \mathbf{Y}_n \Delta_n, \Delta_n)$ be the observations at hand. The missing at random mechanism in considered in the following sense

$$\mathbb{P}(\mathbb{I}_{\{\Delta=1\}} | \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) = \mathbb{P}(\mathbb{I}_{\{\Delta=1\}} | \mathbf{X}_1) := p^0(\mathbf{X}_1).$$

Note that the relation (23) can be written

$$\mathbb{E} \left[\frac{\mathbb{I}_{\{\Delta=1\}}}{p^0(\mathbf{X}_1)} \{ \mathbb{I}_{\{\mathbf{Y}=1, \mathbf{X}_2 \geq f_\theta(\mathbf{X}_1)\}} + \mathbb{I}_{\{\mathbf{Y}=0, \mathbf{X}_2 < f_\theta(\mathbf{X}_1)\}} \} \right].$$

We define

$$\mathbf{m}_{\theta,p}(\mathbf{Z}) = \frac{\mathbb{I}_{\{\Delta=1\}}}{p(\mathbf{X}_1)} \left\{ \mathbb{I}_{\{Y=1, X_2 \geq f_\theta(\mathbf{X}_1)\}} + \mathbb{I}_{\{Y=0, X_2 < f_\theta(\mathbf{X}_1)\}} \right\},$$

here the infinite dimensional nuisance parameter $p(\cdot)$ belonging to some functional space \mathcal{P} to be specified later. Consequently, the estimator θ_n of θ_0 is given by

$$\theta_n = \arg \max_{\theta \in \Theta} \mathbb{P}_n \mathbf{m}_{\theta, \hat{p}},$$

where, for any x and a bandwidth sequence $h = h_n$,

$$\hat{p}(x) = \sum_{i=1}^n \frac{K_h(x - \mathbf{X}_{i1})}{\sum_{j=1}^n K_h(x - \mathbf{X}_{j1})} \mathbb{I}_{\{\Delta_i=1\}},$$

where the kernel function $K(\cdot)$ is assumed to be a density function having support $[-1, 1]$, $K_h(u) = \frac{K(\frac{u}{h})}{h}$. Non parametric regression with missing have long attracted a great deal of attention, for good sources of references to research literature in this area along with statistical applications consult Müller (2009), Pérez-González et al. (2009) and Koul et al. (2012) among many others.

7.3 Binary Choice Model with Missing Data

Let us define the binary choice model, in the linear regression function framework, by

$$\begin{cases} U = \mathbf{X}^\top \boldsymbol{\beta} - \varepsilon, \\ Y = \mathbb{I}(U \geq 0), \end{cases}$$

where we assume that ε is zero median conditionally on \mathbf{X} . The random variable Y is missing at random with the probability, to observe Y , depending on \mathbf{X} via the following relation

$$\mathbb{P}(\mathbb{I}_{\{\Delta=1\}} | \mathbf{X}, Y) = \mathbb{P}(\mathbb{I}_{\{\Delta=1\}} | \mathbf{X}^\top \boldsymbol{\gamma}) := p(\mathbf{X}^\top \boldsymbol{\gamma}),$$

where $\Delta = 1$ when we observe Y and 0 elsewhere. The observed data for the preceding model are given by of i.i.d. triplets $(\mathbf{X}_1, Y_1 \Delta_1, \Delta_n), \dots, (\mathbf{X}_n, Y_n \Delta_n, \Delta_n)$. To estimate $p_\gamma(z) = \mathbb{P}(\mathbb{I}_{\{\Delta=1\}} | \mathbf{X}^\top \boldsymbol{\gamma} = z)$, we use the following

$$\hat{p}_\gamma(z) = \sum_{i=1}^n \frac{K_h(\mathbf{X}_i^\top \boldsymbol{\gamma} - z)}{\sum_{j=1}^n K_h(\mathbf{X}_j^\top \boldsymbol{\gamma} - z)} \mathbb{I}_{\{\Delta_i=1\}}.$$

The parameter estimate is given by

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = \arg \max_{\boldsymbol{\beta}, \boldsymbol{\gamma}} \mathbb{P}_n \mathbf{m}_{\boldsymbol{\beta}, \boldsymbol{\gamma}, \hat{p}_\gamma},$$

where

$$\mathbf{m}_{\boldsymbol{\beta}, \boldsymbol{\gamma}, p} = \frac{\mathbb{I}_{\{\Delta=1\}}}{p(\mathbf{X}_i^\top \boldsymbol{\gamma})} [2 \mathbb{I}_{\{Y_i=1\}} - 1] \mathbb{I}_{\{\mathbf{X}_i^\top \boldsymbol{\beta} \geq 0\}}.$$

The existing theory cannot be applied here by the fact that the function $\mathbf{m}_{\beta,\gamma,p}$ is smooth in γ but non-smooth in β .

Now we will study in full detail the example in Sect. 7.2 and we work out the verification of the conditions of Theorems 3.2, 3.5, 3.8 and 3.9 the most of this conditions verified in Sect. 7 of Delsol and Van Keilegom (2020) by noting that $\nu = 2$ and $\ell \equiv 1$, so our focuses is to verify the conditions needed for the m out of n bootstrapped version. In the beginning we give some information about the nuisance function and her space and some notation. Let \mathcal{P} be the space of functions $p : \mathbf{R}_{\mathbf{X}_1} \rightarrow \mathbb{R}$ that are continuously differentiable, for which

$$\sup_{\mathbf{x}_1 \in \mathbf{R}_{\mathbf{X}_1}} p(\mathbf{x}_1) \leq M < \infty, \sup_{\mathbf{x}_1 \in \mathbf{R}_{\mathbf{X}_1}} |p'(\mathbf{x}_1)| \leq M \text{ and } \inf_{\mathbf{x}_1 \in \mathbf{R}_{\mathbf{X}_1}} p(\mathbf{x}_1) > \eta/2,$$

where

$$\eta = \inf_{\mathbf{x}_1 \in \mathbf{R}_{\mathbf{X}_1}} p^0(\mathbf{x}_1) > 0$$

and $\mathbf{R}_{\mathbf{X}_1}$ is the support of \mathbf{X}_1 , where we suppose it is a compact subspace of \mathbb{R} . We equip the space \mathcal{P} with the supremum norm:

$$d_{\mathcal{P}}(p_1, p_2) = \sup_{\mathbf{x}_1 \in \mathbf{R}_{\mathbf{X}_1}} |p_1(\mathbf{x}_1) - p_2(\mathbf{x}_1)| \text{ for any } p_1, p_2 \in \mathcal{P}.$$

After, the conditions of the consistency are verified as follows, (A1) holds true provided the functions $p_0(\cdot)$ and $K(\cdot)$ are continuously differentiable. For assumption (A2) we can showing that the bracketing number of the class $\mathcal{F} = \{\mathbf{m}_{\theta,p}, \theta \in \Theta, p \in \mathcal{P}\}$; $N_{[\cdot]}(\epsilon, \mathcal{F}, \mathbb{L}_{\mathbb{P}})$ is finite for all $\epsilon > 0$, by using Corollary 2.7.2 of van der Vaart and Wellner (1996), we get

$$N_{[\cdot]}(\epsilon, \mathcal{P}, \mathbb{L}_{\mathbb{P}}) \leq \exp\{\mathfrak{K}\epsilon^{-1}\}, \tag{24}$$

and

$$N_{[\cdot]}(\epsilon, \{f_{\theta}, \theta \in \Theta\}, \mathbb{L}_{\mathbb{P}}) \leq \exp\{\mathfrak{K}\epsilon^{-1}\},$$

by the properties of the set \mathcal{P} and the fact that $\mathbf{x} \mapsto f_{\theta}(\mathbf{x})$ is continuously differentiable over θ with bounded derivative and as a consequence it's easily to show that

$$N_{[\cdot]}(\epsilon, \mathcal{T}, \mathbb{L}_{\mathbb{P}}) \leq \exp\{\mathfrak{K}\epsilon^{-1}\}, \tag{25}$$

for the class $\mathcal{T} = \left\{ (\mathbf{x}_1, \mathbf{x}_2) \rightarrow \Pi_{\{x_2 \geq f_{\theta}(\mathbf{x}_1)\}} : \theta \in \Theta \right\}$. From (24) and (25) we get;

$$N_{[\cdot]}(\epsilon, \mathcal{F}, \mathbb{L}_{\mathbb{P}}) \leq \exp\{\mathfrak{K}\epsilon^{-1}\}.$$

Then assumption (A3) is straightforward. Assumption (A4) is an identifiability condition to ensure the uniqueness of θ_0 and (A5) is verified by construction of the estimator θ_n . The consistency of θ_n is then follows. For the conditions of the bootstrap version they are verified as follows; fist part of assumption (AB1) is satisfied by definition of the m out of n bootstrap, where the second part in this situation follows directly by noting that if $r_n = n^{\kappa}$, we get $r_m = m^{\kappa}$ for some $\kappa > 0$, by consequent we have $r_m^2 = o(r_n^2)$. For (AB2) as mentioned in remark 3.1(v) we take $\hat{p}_m(\cdot) = \hat{p}(\cdot)$ where we replace the variables \mathbf{X}_{1i} and Δ_i by \mathbf{X}_{1i}^* and Δ_i^* respectively in $\hat{p}(\cdot)$; i.e.,

$$\hat{p}_m(x_1) = \sum_{i=1}^m \frac{K_h(x_1 - \mathbf{X}_{i1}^*)}{\sum_{j=1}^m K_h(x_1 - \mathbf{X}_{j1}^*)} \mathbb{I}_{\{\Delta_i^*=1\}},$$

we remark that

$$\mathbb{P}_W \left(\frac{1}{m} \sum_{i=1}^m K_h(x_1 - \mathbf{X}_{i1}^*) \mathbb{I}_{\{\Delta_i^*=1\}} \right) = \frac{1}{n} \sum_{i=1}^n K_h(x_1 - \mathbf{X}_{i1}) \mathbb{I}_{\{\Delta_i=1\}},$$

which implies $d_{\mathcal{H}}(\hat{p}_m, \hat{p}) = o_{\mathbb{P}_W^*}(1)$ i.p. By the triangular inequality we get

$$d_{\mathcal{H}}(\hat{p}_m, p^0) \leq d_{\mathcal{H}}(\hat{p}_m, \hat{p}) + d_{\mathcal{H}}(\hat{p}, p^0) = o_{\mathbb{P}_W^*}(1), \quad \text{i.p.}$$

(AB3) is verified by construction of the estimator θ_m . Which implies the consistency of θ_m . Next for the rate of convergence we show only conditions (B2) and (B3). For (B2), it suffices by **remark 3.3(ii)** to show (4) and (5). For that by uses of the relation between covering and bracketing numbers and Corollary 2.7.2 of van der Vaart and Wellner (1996) we get that

$$\log N \left(\epsilon \left\| M_{\delta, \delta'_1} \right\|_{\mathbb{L}_2(\mathbb{Q})}, \mathcal{M}_{\delta, \delta'_1, \mathbb{L}_2}(\mathbb{Q}) \right) \leq \exp\{ \mathfrak{K} \epsilon^{-1} \},$$

for every probability measure \mathbb{Q} on \mathbb{R}^4 , which implies our relation in (4), (5) is verified by the choice $\varphi(\delta) = \sqrt{\delta}$ as consequence we get (B2). For (B3), it follows directly like in Sect. 7 of the same reference which described this example and by the choice of the two functions $\psi_1(\cdot)$ and $\psi_2(\cdot)$ given in **Remark 3.3(iii)**, which implies (B3). By their discussion for the rates r_n, v_n and the bandwidth h of the kernel; it follows

$$\theta_n - \theta_0 = O_{\mathbb{P}^*}(n^{-1/3}).$$

We verify the assumption (BB1) as in the verification of condition (AB2) by choosing $\hat{p}_m(\cdot) = \hat{p}(\cdot)$ we get $v_m^{-1} = \sqrt{\frac{\log m}{mh}} + h$, where $h = h_m$. Assumption (BB2) holds by the same argument given for (B2). For assumption (BB3), we check conditions (b)-(d) of **Remark 3.3(iii)**. We obtain

$$\Gamma(\theta_0, p) = \mathbb{P} \left[\frac{p_0(\mathbf{X}_1)}{p(\mathbf{X}_1)} \left\{ 1 - 2\mathbb{P}(\mathbb{I}_{\{Y=1\}} | \mathbf{X}_1, \mathbf{X}_2) \right\} f_{\mathbf{X}_2 | \mathbf{X}_1}(f_{\theta_0}(\mathbf{X}_1)) \frac{\partial}{\partial \theta} f_{\theta_0}(\mathbf{X}_1) \right], \quad (26)$$

and

$$\Lambda(\theta_0, p) = \mathbb{P} \left[\frac{p_0(\mathbf{X}_1)}{p(\mathbf{X}_1)} \left\{ 1 - 2\mathbb{P}(\mathbb{I}_{\{Y=1\}} | \mathbf{X}_1, \mathbf{X}_2) \right\} \left\{ f'_{\mathbf{X}_2 | \mathbf{X}_1}(f_{\theta_0}(\mathbf{X}_1)) \left(\frac{\partial}{\partial \theta} f_{\theta_0}(\mathbf{X}_1) \right)^2 + f_{\mathbf{X}_2 | \mathbf{X}_1}(f_{\theta_0}(\mathbf{X}_1)) \frac{\partial^2}{\partial \theta^2} f_{\theta_0}(\mathbf{X}_1) \right\} \right],$$

provided the derivatives in $\Lambda(\theta_0, p)$ all exist. By the definition of maximum it follows that $\Gamma(\theta_0, p^0) = 0$ and $\Lambda(\theta_0, p^0)$ is negative. Noting that

$$\|\Gamma(\theta_0, \hat{p}_m)\| = O_{\mathbb{P}_W}(r_m^{-1}) \text{ i.p.}$$

if r_m satisfies

$$r_m \left(m^{-1/2} + h_m + \frac{\log m}{mh_m} \right) = O(1),$$

by noting that the expectation in (26) is taken with respect to \mathbf{Z} and \mathbf{W} when we are working with \hat{p}_m , since our function are measurable, we obtain such result by applying Fubini’s Theorem. This condition on r_m and the other given in (BB2) which is satisfied for $r_m = O(m^{1/3})$ are reconcilable provided

$$mh_m^3 = O(1) \text{ and } \frac{(\log m)^{3/2}}{mh_m^{3/2}} = O(1).$$

Note that if we assume that $p^0(\cdot)$ is twice continuously differentiable we can weaken the first condition to $mh_m^6 = O(1)$, as a consequence we get the v_m^{-1} of \hat{p}_m would be $O\left(\sqrt{\frac{\log m}{mh_m} + h_m^2}\right)$, which is faster than $r_m^{-1} = m^{-1/3}$ of θ_m provided $mh_m^3 \rightarrow \infty$. The level of complexity of the latter case is less than the case where p^0 is only once differentiable, And we do not discuss it any further, therefore. We conclude that,

$$\theta_m - \theta_0 = O_{\mathbb{P}_W^*}(m^{-1/3}) \text{ i.p.}$$

Finally, for the weak convergence of θ_n , we note that our assumptions (C4) is satisfied for $j_n = \sqrt{n}$ like in Remark3.5 (iii) and (C9) hold similarly to (B2). By consequence $n^{1/3}(\theta_n - \theta_0)$ converge weakly, where assumption (CB1) follows from part (ii) of Theorem 3.5 and condition (BB1), by similar proof of condition (BB2) we get (CB2). We get from Remark 3.3 (iii), (vi) and Remark 3.5 (viii) that assumption (CB3) holds, provided that

$$|\Lambda(\theta_0, p^0)| < \infty.$$

Clearly we have for some positive constant $c > 0$ that $m^{-1/3} < C$. For assumption (CB4), we have

$$r_m W_m(\gamma) = r_m \Gamma(\theta_0, \hat{p}_m)\gamma = o_{\mathbb{P}_W}(1) \text{ i.p.,}$$

provided $mh_m^3 = o(1)$ and $\frac{\log^{3/2} m}{mh_m^{3/2}} = o(m^{-1/2})$, by using what we discuss already for (BB3). Next, by the result given to the process in (16) i.e., the process $\gamma \mapsto \mathbb{G}_n \frac{r_m^2}{\sqrt{m}} \tilde{\mathbf{m}}_{r_m, h^0}$ converges weakly to the process $\mathbb{G}(\gamma)$ and condition (AB1), we get

$$\begin{aligned} r_m^2 \hat{\mathbb{P}}_m \tilde{\mathbf{m}}_{r_m, p^0} &= r_m^2 \left[\left(\hat{\mathbb{P}}_m - \mathbb{P}_n \right) \tilde{\mathbf{m}}_{r_m, p^0} + \sqrt{\frac{m}{n}} \mathbb{G}_n \frac{\tilde{\mathbf{m}}_{r_m, p^0}}{\sqrt{m}} + \mathbb{P} \tilde{\mathbf{m}}_{r_m, p^0} \right] \\ &= r_m^2 \left(\hat{\mathbb{P}}_m - \mathbb{P}_n \right) \tilde{\mathbf{m}}_{r_m, p^0} + \frac{1}{2} \Lambda(\theta_0, p^0) \gamma^2 + o_{\mathbb{P}}(1), \end{aligned}$$

with $\Gamma(\theta_0, p^0) = 0$ and

$$\Lambda(\gamma) = \frac{1}{2}\Lambda(\theta_0, p^0)\gamma^2.$$

The process $\gamma \rightarrow r_m^2 \left(\hat{\mathbb{P}}_m - \mathbb{P}_n \right) \tilde{\mathbf{m}}_{\frac{\gamma}{r_m}, p^0}$ are the same given in Lee (2012) where there is no presence of nuisance parameter. Hence, we can follow the same steps given in Lemma 1 of Lee (2012) and get the convergence of the marginals using Lindeberg's condition and some regularity assumption on f_{X_1/X_2} and $\theta \mapsto f_\theta$. By construction of the estimator θ_m , condition (CB5) follows. Then we get the asymptotic distribution of $r_m(\theta_m - \theta_n)$ from part (ii) of Theorem 3.9.

Acknowledgements The authors are indebted to the Editor-in-Chief, Associate Editor and the referee for their very valuable comments, suggestions careful reading of the article which led to a considerable improvement of the manuscript. The Third author gratefully acknowledges the funding received towards his PhD from the Algerian government PhD fellowship.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

References

- Alin A, Martin MA, Beyaztas U, Pathak PK (2017) Sufficient m -out-of- $n(m/n)$ bootstrap. *J Stat Comput Simul* 87(9):1742–1753
- Allaire G (2005) Analyse numérique et optimisation: une introduction à la modélisation mathématique et à la simulation numérique. Editions Ecole (Polytechnique)
- Alvarez-Andrade S, Bouzebda S (2013) Strong approximations for weighted bootstrap of empirical and quantile processes with applications. *Stat Methodol* 11:36–52
- Alvarez-Andrade S, Bouzebda S (2015) On the local time of the weighted bootstrap and compound empirical processes. *Stoch Anal Appl* 33(4):609–629
- Alvarez-Andrade S, Bouzebda S (2019) Some selected topics for the bootstrap of the empirical and quantile processes. *Theory Stoch Process* 24(1):19–48
- Arcones MA, Giné E (1992) On the bootstrap of M -estimators and other statistical functionals. In *Exploring the limits of bootstrap* (East Lansing, MI, 1990), Wiley Ser Probab Math Statist Probab. Math. Statist., pages 13–47. Wiley, New York
- Bickel PJ, Sakov A (2008) On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statist Sinica* 18(3):967–985
- Bickel PJ, Klaassen CAJ, Ritov Y, Wellner JA (1993) Efficient and adaptive estimation for semiparametric models. Johns Hopkins Series in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD
- Bickel PJ, Götzte F, van Zwet WR (1997) Resampling fewer than n observations: gains, losses, and remedies for losses. *Statist Sinica* 7(1), 1–31. Empirical Bayes, sequential analysis and related topics in statistics and probability (New Brunswick, NJ, 1995)
- Bose A, Chatterjee S (2001) Generalised bootstrap in non-regular M -estimation problems. *Statist Probab Lett* 55(3):319–328
- Bouzebda S (2010) Bootstrap de l'estimateur de Hill: théorèmes limites. *Ann ISUP* 54(1-2), 61–72
- Bouzebda S, Linnios N (2013) On general bootstrap of empirical estimator of a semi-Markov kernel with applications. *J Multivariate Anal* 116:52–62
- Bouzebda S, Papamichail C, Linnios N (2018) On a multidimensional general bootstrap for empirical estimator of continuous-time semi-Markov kernels with applications. *J Nonparametr Stat* 30(1):49–86
- Chen X, Linton O, Van Keilegom I (2003) Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* 71(5):1591–1608
- Cheng G, Huang JZ (2010) Bootstrap consistency for general semiparametric M -estimation. *Ann Statist* 38(5):2884–2915
- Chernick MR (2008) *Bootstrap methods: a guide for practitioners and researchers*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition

- Datta S, McCormick WP (1995) Bootstrap inference for a first-order autoregression with positive innovations. *J Amer Statist Assoc* 90(432):1289–1300
- Davison AC, Hinkley DV (1997) Bootstrap methods and their application, volume 1 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge. With 1 IBM-PC floppy disk (3.5 inch; HD)
- Delsol L, Van Keilegom I (2020) Semiparametric M -estimation with non-smooth criterion functions. *Ann Inst Statist Math* 72(2):577–605
- Dudley RM (1999) Uniform central limit theorems, volume 63 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge
- Efron B (1979) Bootstrap methods: another look at the jackknife. *Ann Statist* 7(1):1–26
- Efron B, Tibshirani RJ (1993) An introduction to the bootstrap, volume 57 of Monographs on Statistics and Applied Probability. Chapman and Hall, New York
- El Bantli F (2004) M -estimation in linear models under nonstandard conditions. *J Statist Plann Inference* 121(2):231–248
- Giné E, Zinn J (1989) Necessary conditions for the bootstrap of the mean. *Ann Statist* 17(2):684–691
- Götze F, Račkauskas A (2001) Adaptive choice of bootstrap sample sizes. In *State of the art in probability and statistics (Leiden, 1999)*, volume 36 of IMS Lecture Notes Monogr. Ser., pages 286–309. Inst. Math. Statist., Beachwood, OH
- Hall P (1992) The bootstrap and Edgeworth expansion. Springer Series in Statistics. Springer-Verlag, New York
- Hall P, Horowitz JL, Jing B-Y (1995) On blocking rules for the bootstrap with dependent data. *Biometrika* 82(3):561–574
- Hoffmann-Jørgensen J (1991) Stochastic processes on Polish spaces. Various publications series. Aarhus Universitet, Matematisk Institut
- Ichimura H (1993) Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J Econometrics* 58(1–2):71–120
- Kim J, Pollard D (1990) Cube root asymptotics. *Ann Statist* 18(1):191–219
- Kosorok MR (2008) Introduction to empirical processes and semiparametric inference. Springer Series in Statistics, Springer, New York
- Koul HL, Müller UU, Schick A et al (2012) The transfer principle: a tool for complete case analysis. *Ann Stat* 40(6):3031–3049
- Kristensen D, Salanié B (2017) Higher-order properties of approximate estimators. *J Econometrics* 198(2):189–208
- Lahiri SN (1992) On bootstrapping M -estimators. *Sankhyā Ser A* 54(2):157–170
- Lee SMS (2012) General M -estimation and its bootstrap. *J Korean Statist Soc* 41(4):471–490
- Lee SMS, Pun MC (2006) On m out of n bootstrapping for nonstandard M -estimation with nuisance parameters. *J Amer Statist Assoc* 101(475):1185–1197
- Lee SMS, Yang P (2020) Bootstrap confidence regions based on M -estimators under nonstandard conditions. *Ann Statist* 48(1):274–299
- Ma S, Kosorok MR (2005) Robust semiparametric m -estimation and the weighted bootstrap. *J Multivar Anal* 96(1):190–217
- Müller UU et al (2009) Estimating linear functionals in nonlinear regression with responses missing at random. *Ann Stat* 37(5A):2245–2277
- Pakes A, Olley S (1995) A limit theorem for a smooth class of semiparametric estimators. *J. Econometrics* 65(1):295–332
- Pakes A, Pollard D (1989) Simulation and the asymptotics of optimization estimators. *Econometrica* 57(5):1027–1057
- Pérez-González A, Vilar-Fernández JM, González-Manteiga W (2009) Asymptotic properties of local polynomial regression with missing data and correlated errors. *Ann Inst Stat Math* 61(1):85–109
- Pfanzagl J (1990) Estimation in semiparametric models, vol 63. Lecture Notes in Statistics. Springer-Verlag, New York. Some recent developments
- Politis DN, Romano JP, Wolf M (1999) Subsampling. Springer Series in Statistics. Springer-Verlag, New York
- Pollard D (1985) New ways to prove central limit theorems. *Economet Theor* 1(3):295–313
- Shao J, Tu DS (1995) The jackknife and bootstrap. Springer Series in Statistics. Springer-Verlag, New York
- Swanepoel JWH (1986) A note on proving that the (modified) bootstrap works. *Comm Statist A-Theory Methods* 15(11):3193–3203
- van de Geer SA (2000) Applications of empirical process theory, volume 6 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge

- van der Vaart AW, Wellner JA (1996) Weak convergence and empirical processes. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics
- Wei B, Lee SMS, Wu X (2016) Stochastically optimal bootstrap sample size for shrinkage-type statistics. *Stat Comput* 26(1–2):249–262
- Wellner JA, Zhan Y (1996) Bootstrapping Z-estimators. Preprint
- Zhan Y (2002) Central limit theorems for functional Z-estimators. *Statist. Sinica* 12(2):609–634

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