Asymptotic and bootstrap properties of rank regressions

Subbotin, Viktor
Northwestern University

08. November 2007

Asymptotic and Bootstrap Properties of Rank Regressions

Viktor Subbotin*
Northwestern University

March 20, 2008

Abstract

The paper develops the bootstrap theory and extends the asymptotic theory of rank estimators, such as the Maximum Rank Correlation Estimator (MRC) of Han (1987), Monotone Rank Estimator (MR) of Cavanagh and Sherman (1998) or Pairwise-Difference Rank Estimators (PDR) of Abrevaya (2003). It is known that under general conditions these estimators have asymptotic normal distributions, but the asymptotic variances are difficult to find. Here we prove that the quantiles and the variances of the asymptotic distributions can be consistently estimated by the nonparametric bootstrap. We investigate the accuracy of inference based on the asymptotic approximation and the bootstrap, and provide bounds on the associated error. In the case of MRC and MR, the bound is a function of the sample size of order close to $n^{-1/6}$. The PDR estimators belong to a special subclass of rank estimators for which the bound is vanishing with the rate close to $n^{-1/2}$. The theoretical findings are illustrated with Monte-Carlo experiments and a real data example.

JEL Classifications: C12, C14

Keywords: Rank Estimators, Bootstrap, M-Estimators, U-Statistics, U-Processes

*Corresponding information: Viktor Subbotin, Northwestern University, Economics Department, 2001 Sheridan Road, Office 302, Evanston, Illinois 60208. Phone: 630-205-4464. E-mail: v-soubbotine@northwestern.edu.

1 Introduction

Several semiparametric estimators recently introduced in the econometrics literature are based on the rank correlation between the dependent and explanatory variables. The first was suggested by Han (1987), and is called the Maximum Rank Correlation estimator (MRC). It applies to the generalized regression model, given by the relation

$$Y = D \circ F\left(X'\beta_0, \varepsilon\right). \tag{1}$$

Here X and ε are independent random variables, the function D is nondecreasing, F is strictly increasing in both arguments, and β_0 is a finite-dimensional vector of parameters. Binary choice models, censored regression models, the Box and Cox transformation models, and proportional and additive hazard models are particular examples of (1). In general, the vector β_0 is identified up to scale even if D, F, and the distribution of the error term are not specified. Given a sample (X_i, Y_i), $i = 1, ..., n$, of i.i.d. observations, β_0 can be consistently estimated by maximizing the rank correlation objective function

$$\sum_{i \neq j} \mathbf{1}\{Y_i > Y_j\}\mathbf{1}\{X_i'\beta > X_j'\beta\}. \tag{2}$$

with respect to β under a scale normalization restriction.

Another example is the Monotone Rank Estimator (MR) of Cavanagh and Sherman (1998). Here β_0 is identified (up to scale) by the relation

$$E\left[S\left(Y\right)|X\right] = f\left(X'\beta_0\right)$$

where f is an increasing, nonconstant function, and $S\left(Y_i\right)$ is either an increasing function of Y_i, or the sample rank of Y_i (i.e. $S\left(Y_i\right) = \sum_k \mathbf{1}\left\{Y_i > Y_k\right\}$). In the first case, MR is a maximization estimator with the objective function

$$\sum_{i \neq j} S\left(Y_i\right)\mathbf{1}\{X_i'\beta > X_j'\beta\}. \tag{3}$$

In the second, the objective function is

$$\sum_{i,j,k \text{ distinct}} \mathbf{1}\left\{Y_i > Y_k\right\}\mathbf{1}\{X_i'\beta > X_j'\beta\}. \tag{4}$$

Abrevaya (2003) considered a special case of the generalized regression model, the transformation model

$$f(Y) = X'\beta_0 + \varepsilon \tag{5}$$

where f is a *strictly* increasing unknown function, and X, ε and β_0 are as above. He proposed two pairwise-difference rank estimators of β_0. The PDR3 estimator maximizes the objective function

$$\sum_{i,j,k \text{ distinct}} \left(\mathbf{1}\{Y_i > Y_j\} - \mathbf{1}\{Y_j > Y_k\}\right) \mathbf{1}\left\{(X_i - X_j)'\beta > (X_j - X_k)'\beta\right\},$$

and the PDR4 estimator maximizes the objective function

$$\sum_{i,j,k,l \text{ distinct}} \left(\mathbf{1}\{Y_i > Y_j\} - \mathbf{1}\{Y_k > Y_l\}\right) \mathbf{1}\left\{(X_i - X_j)'\beta > (X_k - X_l)'\beta\right\}.$$

Estimators with similar structure have been proposed for such models as the transformation model with observed or unobserved truncation (Abrevaya (1999b), Khan and Tamer (2007)), the binary response model with panel data (Lee (1999)) and a version of the generalized regression model for panel data with fixed-effects (Abrevaya (1999a)), among others. See also the papers by Han (1987b), Asparouhova et al. (2002), and Chen (2002), who considered a parametric and a nonparametric estimation of the link function f in the transformation model using rank regression techniques.

Rank correlation estimators have several advantages. First, they rely on relatively weak identification assumptions. Second, as shown by Sherman (1993), they are root-n-consistent and asymptotically normal. Most importantly, they do not require a choice of any tuning parameters (bandwidths, etc.), unlike all other presently known asymptotically normal semiparametric estimators (such as the average derivative method of Powell, Stock and Stoker (1989), the semiparametric least-squares estimator of Ichimura (1993), the sieve minimum-distance estimator of Ai and Chen (2003) or the semiparametric maximum likelihood method (for binary response models) of Klein and Spady (1993)). This last property is useful for empirical work, as choosing bandwidths or other tuning parameters is not always easy in practice.

Finally, rank estimators can be applied when the distribution of ε has heavy tails. The estimation of β in the transformation model (5) by the method of least-squares, for a known f, or the nonparametric methods minimizing a quadratic distance, e.g. estimators of Ai and Chen or Ichimura, for an unknown f, is inconsistent if the second moment of ε is infinite. In practical terms this means that these estimators can be sensitive to outliers. Estimators like MRC or PDR, on the other hand, are root-n-consistent and asymptotically normal even if the first absolute moment of ε is infinite.

This paper is concerned with inference in rank regressions. Test statistics, their critical values, and confidence intervals for components of β_0 can be constructed in the usual way based on the limiting normal distribution. Unfortunately, the variance of that distribution depends on moments of random variables that are not directly observed (first and second-order derivatives of certain conditional expectations), and special procedures are needed for its estimation. Two methods that are available at present are the numerical derivative method of Pakes and Pollard (1989), and a nonparametric method of Sherman (1993) and Cavanagh and Sherman (1998). However, both have drawbacks. First, they depend on tuning parameters (step sizes for numerical differentiation or bandwidths for kernel regressions), which deprives rank estimators of their primary appeal. No objective, data-driven mechanism has been developed to set these parameters in practical applications. The numerical derivative method involves a finite-difference approximation of the second-order derivatives and can produce unstable results. The nonparametric method, which avoids the estimation of the second-order derivatives, requires considerable programming effort, as the analytical expressions for the variances that it uses are specific for each particular estimator, and are sometimes very complicated (as in the case of PDR3, for example). Finally, the nonparametric method is numerically intensive in large samples, with the computational burden rising with the sample size as $O\left(n^2\right)$ for MRC or MR, and $O\left(n^4\right)$ for PDR4[1].

Alternatively, the asymptotic distribution can be estimated by resampling methods, particularly, the nonparametric bootstrap of Efron (1979). This approach is free of tuning parameters, and is easy to implement. Unlike in most other econometric settings, the bootstrap of rank correlation estimators can be less computationally demanding than direct variance estimation, due to availability of fast algorithms for evaluating their objective functions. For

[1] These and the following estimates of the computational complexity assume that the full sample is used for inference. When n is large, inference can be performed, at the expense of lower precision, using a randomly chosen subsample of data.

example, one evaluation of the objective function can be reduced to only $O(n \log n)$ operations for MR (Cavanagh and Sherman (1998)), and MRC (Abrevaya (1999b)), and to $O(n^2 \log n)$ operations for both PDR (Abrevaya (2003)). The same efficient algorithms can be used in the nonparametric bootstrap, making it feasible and often more attractive computationally than other alternatives.

Prior to our work, it has not been known if the nonparametric bootstrap consistently estimates the asymptotic distribution of rank estimators. The fact that an estimator is root-n-consistent and asymptotically normal does not guarantee consistency of the bootstrap (see Abadie and Imbens (2006) for a counterexample). Here we prove its consistency for rank regressions. The regularity conditions that we require are, up to a minor qualification, the same as the assumptions under which Sherman's asymptotic normality result holds.

Next, we investigate the accuracy of inference in rank regressions. The normal distribution (with the true or estimated asymptotic variance) and the bootstrap distribution approximate the finite-sample distribution of rank estimators with an error. While the error converges to zero as n grows (under conditions of Sherman's theorem and our bootstrap consistency theorem), it may do so slowly. In practice, this means that the confidence intervals and tests of hypotheses constructed using either approximation may have coverage probabilities and levels very different from the nominal ones when the sample size is finite.

The problem of the accuracy of inference is well understood in the case of an estimator that is a smooth function of sample moments (see e.g. Bhattacharya and Rao (1976) and Hall (1992)). Then, confidence intervals based on the normal approximation typically attain the desired coverage probability up to an error of order $O\left(n^{-1/2}\right)$ for one-sided confidence intervals and $O\left(n^{-1}\right)$ for two-sided symmetric confidence intervals. In the case of M-estimators with non-smooth criterion functions, the exact order of the approximation error is known only in several special cases (such as the Least Absolute Deviation estimator studied by De Angeles et al. (1993)). Some results are available for nonparametric methods that are applicable to models (1) or (5). Nishiyama and Robinson (2005) studied the accuracy of inference for the normal and the bootstrap approximations for the average derivative estimator and showed that it can achieve the same degree of accuracy as in parametric methods. However, this conclusion relies on restrictive moment and smoothness conditions. Particularly, there is a hidden "curse-of-dimensionality" effect: the conditional expectation $E[Y|X'\beta]$ has to have progressively higher numbers of bounded derivatives in the single index $X'\beta$ as the dimension of X grows, and progressively higher orders of kernels have to be used in associated nonparametric regressions[2].

Below we give an upper bound on the error of approximation of the finite-sample distributions of MRC, MR, and the other rank estimators listed above. The bound is the same for approximations by both the bootstrap distribution and the normal distribution with the true variance. In the case of MRC, the error converges to zero with the rate arbitrarily close to $n^{-1/6}$. The rate is slower for the MR estimator if the function S is not bounded, but it also approaches the order of $n^{-1/6}$ if S has sufficiently many finite moments. The result holds under mild regularity conditions and is not subject to the curse of dimensionality. We further show that, under somewhat stronger assumptions, the PDR3 and PDR4 estimators have a much smaller approximation error, close to $n^{-1/2}$ in the case of PDR3 and exactly $n^{-1/2}$ in the case of PDR4. Therefore, in one-sided tests and confidence intervals, the pairwise-difference rank estimators achieve the same order of accuracy as the classical parametric estimators. We are not aware of existence of smoothing-based nonparametric techniques applicable to model (5) that would achieve this degree of precision of inference under the same regularity conditions.

Our work has a close connection to the statistical literature on the asymptotics and the bootstrap of U-processes. However, we could not find any previous work on the bootstrap of maximizers of general U-processes. Here we develop such theory, and apply it to the rank estimators.

The paper is organized as follows. Section 2 presents the asymptotic and bootstrap theory of rank estimators. Sections 3 illustrates the theory with Monte-Carlo experiments and a real-data example. The proofs are developed in Appendix. Section 4 concludes.

---

[2] The same is true for conditions under which the average derivative estimator is root-n-consistent. The estimator by Ai and Chen also has a hidden curse of dimensionality, since it requires progressively stronger smoothness properties of the unknown functions when the dimension of the vector X grows. Other methods (e.g. Ichimura's estimator) may not have this problem. Unfortunately, the second-order asymptotic properties of Ichimura's estimator are not known. It is likely though that strong smoothness assumptions will be needed for it to have the rate of convergence of order $O\left(n^{-1/2}\right)$ for the error between the finite-sample distribution of the estimator and the asymptotic normal distribution.

2 Asymptotic Theory of Rank Estimators

2.1 The Class of Estimators and Asymptotic Normality Theorems

We first define a class of estimators that includes all of the rank estimators listed in Introduction. The following notation is used: $\mathcal{Z}$ is a vector space, and P is a probability measure on $\mathcal{Z}$;

$$\mathcal{H} = \{h_\theta(z_1, ..., z_m) : \theta \in \Theta \subset \mathbb{R}^d\}$$

is a family of real-valued functions defined on $\mathcal{Z}^m = \mathcal{Z} \times ... \times \mathcal{Z}$ ($m \geq 2$ times), indexed by a vector of parameters θ. It will be a matter of notational convenience to assume that all functions h are symmetric in their z arguments:

$$h_\theta(z_1, ..., z, ..., z', ..., z_m) = h_\theta(z_1, ..., z', ..., z, ..., z_m).$$

Write $P^{m-k}h$, $k = 0, ..., m$, for the partial integral, relative to P, over the last $m - k$ arguments of h:

$$\left(P^{m-k}h\right)(z_1, ..., z_k) = \int h\left(z_1, ..., z_k, Z_{k+1}, ..., Z_m\right) dP\left(Z_{k+1}\right) ... dP\left(Z_m\right)$$

(in particular, $P^0 h = h$).

Assume that the parameter of interest, θ_0, is a global maximum on Θ of the expected value of h_θ, $P^m h_\theta$. Given an i.i.d. sample of observations, $\{Z_1, ..., Z_n\}$, from $(\mathcal{Z}, P)$, one can construct a sample analog of $P^m h_\theta$, a U-process of order m indexed by θ:

$$G_{n,\theta} = U_n^{(m)} h_\theta \equiv \frac{(n-m)!}{n!} \sum_{i_1, ..., i_m, \text{ distinct}} h_\theta(Z_{i_1}, ..., Z_m) \tag{6}$$

(a U-process considered for a specific θ is called a U-statistic. See e.g. Serfling (1980) on the basic properties of U-statistics).

The parameter θ_0 can be estimated by an approximate solution of the sample analog of the population problem:

$$G_{n,\theta_n} \geq \sup_{\theta \in \Theta} \left[G_{n,\theta} - r_{n,\theta}\right], \tag{7}$$

where the remainder term $r_{n,\theta}$ is introduced to ensure measurability of θ_n as in Pakes and Pollard (1989) and may also represent the terms that do not have the structure studied below (e.g. the numerical error of solving the maximization problem)[3].

Under general conditions, θ_n is root-n-consistent for θ_0 and asymptotically normal. Namely, let the following assumptions hold.

Assumption 1 *Θ is a compact set; $P^m h_\theta$, $m \geq 2$, is continuous on Θ and θ_0 is its unique global maximum on Θ.*

Assumption 2 *$\mathcal{H}$ is a Euclidean class[4] of symmetric functions for a P^m-square-integrable envelope H (H is called an envelope for the class $\mathcal{H}$ if $|h| \leq H$ for each $h \in \mathcal{H}$).*

Assumption 3 *Define $\tau_\theta(z) = \left(P^{m-1}h_\theta\right)(z)$. There is an open neighborhood $\mathcal{N} \subset \Theta$ of θ_0 such that*
(i) All mixed partial derivatives of $\tau_\theta(z)$ with respect to θ of orders 1 and 2 exist on $\mathcal{N}$.
(ii) There is a P-integrable function $M(z)$ such that for all z and all θ in $\mathcal{N}$,

$$\left\|\partial^2\tau_\theta(z) - \partial^2\tau_{\theta_0}(z)\right\| \leq M(z)\left\|\theta - \theta_0\right\|,$$

where $\partial^2\tau_\theta$ is the Hessian matrix of τ with respect to θ.
(iii) The gradient of $\tau_\theta(z)$ with respect to θ at θ_0, $\partial\tau_{\theta_0}(z)$, has finite variance relative to P.
(iv) The matrix $A = -P\left[\partial^2\tau_{\theta_0}\right]$ is finite and positive definite.

Assumption 4 *As $\theta \to \theta_0$, $P^2\left[\left(P^{m-2}h_\theta - P^{m-2}h_{\theta_0}\right)^2\right] \to 0$.*

These assumptions are a stylized version of assumptions of Sherman (1993). Assumption 1 is standard for identification. The second assumption says that the class of functions over which maximization is performed is not too big, which is necessary for consistency. Assumptions 3 and 4 repeat the continuity and smoothness conditions of Sherman for asymptotic normality.

For example, in the case of MRC, let $\beta = (\theta, 1)' \in \mathbb{R}^{d+1}$ (to fix the scale, the last component of β is set to 1). The function h_θ is a symmetrized

[3] Below we describe admissible orders of magnitude of $r_{n,\theta}$.
[4] See Appendix, Section 5.3, for the definition and basic properties of Euclidean classes.

version of the kernel in (2) (note that symmetrization does not change the optimization problem):

$$\begin{aligned} h_\theta(z_1, z_2) &= \\ \mathbf{1}\{y_1 &> y_2\}\mathbf{1}\{(\theta', 1)(x_1 - x_2) > 0\} \\ +\mathbf{1}\{y_2 &> y_1\}\mathbf{1}\{(\theta', 1)(x_2 - x_1) > 0\} \end{aligned} \tag{8}$$

where $z = (y, x)$. Han (1987) provided primitive conditions under which h_θ satisfies Assumption 1. Sherman (1993) verified that for a compact Θ, the class of functions $\{h_\theta(z_1, z_2)\}$ is Euclidean for the envelope $H = 1$, and gave conditions on the primitives of the model (1) under which Assumptions 3 and 4 are satisfied. In particular, Assumption 4 will be satisfied if the last component of vector X, denoted x^{d+1}, is continuously distributed conditionally on the vector of the first d components, $\mathcal{X}$. Also, the following condition is sufficient for parts (i)-(iii) of Assumption 3: x^{d+1} is continuously distributed conditionally on $\mathcal{X}$ and Y; the conditional density, $\phi_{x^{d+1}|\mathcal{X},Y}$, is three times differentiable in x^{d+1} for almost all $\mathcal{X}$ and Y and is uniformly bounded together with its derivatives up to order three; and $P\left\|\mathcal{X}\right\|^3 < \infty$[5]. Below, we will refer to these, or similar, sufficient conditions repeatedly. Assumptions 1-4 were verified for the other rank estimators in the corresponding papers listed in Introduction. It is worth noting, however, that Assumptions 1-4 do not rely on the specific structure of rank estimators, but rather on the fact that they all maximize a U-process with sufficiently smooth leading terms. The applicability of our theoretical results, therefore, extends beyond the scope of rank estimators.

Theorem 1, which is essentially due to Sherman (1993) and Arcones, Giné and Chen (1994), says that the estimator θ_n, after a proper normalization and recentering, converges in distribution, uniformly, to a normal law.

Theorem 1 *Let Assumptions 1-4 hold, and* $\sup_{\theta\in\Theta}|r_{n,\theta}| = o_p\left(n^{-1}\right)$. *Define* $\Gamma = m^2A^{-1}Var\left(\partial\tau_{\theta_0}\right)A^{-1}$. *Then θ_n is consistent for θ_0 in probability, and*

$$\sup_{A\in\mathcal{A}}\left|\int_A dF_{n^{1/2}(\theta_n-\theta_0)} - \int_A d\Phi_\Gamma\right| = o\left(1\right) \tag{9}$$

where $F_{n^{1/2}(\theta_n-\theta_0)}$ is the c.d.f. of the random vector $n^{1/2}\left(\theta_n - \theta_0\right)$, Φ_Γ is the c.d.f. of the normal distribution with mean zero and variance Γ, and $\mathcal{A}$ is the collection of all measurable convex sets in $\mathbb{R}^d$.

[5]Sherman assumed that $\mathcal{X}$ has bounded support, but $P\left\|\mathcal{X}\right\|^3 < \infty$ suffices.

To use the result of Theorem 1 for inference, one needs an estimate of the asymptotic variance Γ. The latter, however, depends on moments of the derivatives of the unknown function τ. As explained in Introduction, estimation of these moments may be difficult in practice.

Alternatively, the limiting distribution can be estimated by the non-parametric bootstrap of Efron (1979). Specifically, let $\left\{\hat{Z}_1, ..., \hat{Z}_n\right\}$ be the bootstrap sample, i.e. a collection of independent draws, with replacement, from the sample $\{Z_1, ..., Z_n\}$. The bootstrapped objective function $\hat{U}_n^{(m)} h_\theta$ is formed as in (6) using $\hat{Z}_i$ instead of Z_i:

$$\hat{G}_{n,\theta} = \hat{U}_n^{(m)} h_\theta \equiv \frac{(n-m)!}{n!} \sum_{i_1,...,i_m,\text{ distinct}} h_\theta(\hat{Z}_{i_1}, ..., \hat{Z}_m) \tag{10}$$

The bootstrapped estimator, $\hat{\theta}_n$, is an approximate solution to the corresponding maximization problem:

$$\hat{G}_{n,\hat{\theta}_n} \geq \sup_{\theta \in \Theta} \left[\hat{G}_{n,\theta} - \hat{r}_{n,\theta}\right] \tag{11}$$

with some remainder $\hat{r}_{n,\theta}$.

To prove consistency of the bootstrap, we make one more assumption. It arises because the bootstrap draws, unlike the sample observations, are statistically dependent unconditionally. Note that Assumptions 1-4 provide no bounds on moments of function h if its arguments are statistically dependent. The form of dependency that needs to be explicitly controlled in the bootstrap is that of drawing the same sample realization of vector Z two or more times. To state the assumption formally, define the function

$$H_{\omega_m}(z_1, ..., z_m) = H\left(z_{\omega_m(1)}, ..., z_{\omega_m(m)}\right),$$

where ω_m is a permutation, with repetitions, of numbers $\{1, ..., m\}$; and the function

$$h_\theta^{[m-2]}(z_1, ..., z_{m-2}) = \int h_\theta(z_1, ..., z_{m-2}, Z_m, Z_m)\, dP(Z_m).$$

Assumption 5 *(a) For all ω_m, $P^m H_{\omega_m}^2 < \infty$.*
(b) As $\theta \to \theta_0$, $P^{m-2} h_\theta^{[m-2]} - P^{m-2} h_{\theta_0}^{[m-2]} \to 0$.

Assumption 5 is not restrictive for rank estimators. The moment condition on H_{ω_m} is usually immediate. It is trivially satisfied for bounded functions h (which is the case for the majority of rank estimators). The MR estimator is an example when h may be unbounded, however, the condition $PS^2 < \infty$, required by Assumption 2, also entails the moment condition in Assumption 5 (a). The continuity condition on $h_\theta^{[m-2]}$ is also not difficult to verify. For MRC, for example, it is satisfied vacuously, because in this case $h_\theta^{[m-2]} \equiv 0$. For other estimators, e.g. pairwise-difference rank estimators, $h_\theta^{[m-2]} \neq 0$. However, Assumption 5, similarly to Assumption 4, holds if the last component of the vector of regressors, x^{d+1}, is distributed continuously conditionally on the first d components.

We now give two results showing consistency of the bootstrap. The distribution of the test statistic, $n^{1/2}(\theta_n - \theta_0)$, can be approximated by the conditional (on the data sample) distribution of the bootstrapped statistic, $n^{1/2}(\widehat{\theta}_n - \theta_n)$, or by the normal c.d.f. with zero mean and the conditional variance of $n^{1/2}(\widehat{\theta}_n - \theta_n)$. Both approaches are consistent, although the second relies on slightly stronger regularity conditions.

Theorem 2 *Let the assumptions of Theorem 1 and Assumption 5 hold, and assume that* $\sup_\theta |\hat{r}_{n,\theta}| = o_p(n^{-1})$. *Then the bootstrap estimator of the asymptotic distribution of* $n^{1/2}(\theta_n - \theta_0)$ *is consistent in probability:*

$$\sup_{A \in \mathcal{A}} \left| \int_A d\hat{F}_{n^{1/2}\left(\hat{\theta}_n - \theta_n\right)} - \int_A d\Phi_\Gamma \right| = o_p(1), \tag{12}$$

where $\hat{F}$ *is the conditional c.d.f. of the bootstrapped estimator.*

Theorem 3 *Let the assumptions of Theorem 1 hold, and, additionally,* $P^m H^p < \infty$, $PM^p < \infty$, $P\left\|\partial^2 \tau_{\theta_0}\right\|^p < \infty$, *for a* $p > 2$, *and* $P \sup_\theta |r_{n,\theta}|^2 = o(n^{-1})$. *Then*

$$Var\left[n^{1/2}(\theta_n - \theta_0)\right] \to \Gamma.$$

If also $P^m H_{\omega_m}^p < \infty$ *for each* ω_m, $P \sup_\theta |\hat{r}_{n,\theta}|^2 = o(n^{-1})$, *and Assumption 5 (b) holds, then the bootstrap estimator of the asymptotic variance of* $n^{1/2}(\theta_n - \theta_0)$ *is consistent in probability:*

$$\widehat{Var}\left[n^{1/2}(\widehat{\theta}_n - \theta_n)\right] \to^p \Gamma.$$

Here Var *is the finite sample variance and* $\widehat{Var}$ *is the bootstrap variance conditional on the sample.*

2.2 Rates of Convergence: General Case

Theorems 1 and 2 do not provide insight on how well the bootstrap or the normal approximations of the finite-sample distributions of rank estimators may perform in practice, when the sample size is finite. Next we give bounds on the approximation errors in (9) and (12). Here we consider the rank estimators in general, and in the next subsection we deal with the special case of the pairwise-difference rank estimators.

The bounds that we find are closely related to the continuity properties of the quantity appearing in Assumption 4. In the general case, we impose the following condition.

Assumption 6 *There exist numbers $\delta, C > 0$ such that for all θ_1, θ_2 in the δ-neighborhood of θ_0,*

$$P^2\left[\left(P^{m-2}h_{\theta_1} - P^{m-2}h_{\theta_2}\right)^2\right] \leq C\left\|\theta_1 - \theta_2\right\|. \tag{13}$$

Note that for differentiable kernels h one would normally have, by a Taylor expansion argument, that

$$P^2\left[\left(P^{m-2}h_{\theta_1} - P^{m-2}h_{\theta_2}\right)^2\right] = O\left(\left\|\theta_1 - \theta_2\right\|^2\right). \tag{14}$$

Assumption 6, therefore, reflects a degree of nonsmoothness associated with the criterion function of rank estimators. In fact, for estimators like MRC or MR, both (13), and its reverse:

$$P^2\left[\left(P^{m-2}h_{\theta_1} - P^{m-2}h_{\theta_2}\right)^2\right] \geq c\left\|\theta_1 - \theta_2\right\|. \tag{15}$$

(for a constant $c > 0$) are generally true.

Consider, for example, MRC. One can see that

$$\left[h_{\theta_1}(z_1, z_2) - h_{\theta_2}(z_1, z_2)\right]^2 = \left|h_{\theta_1}(z_1, z_2) - h_{\theta_2}(z_1, z_2)\right|$$

(this is a consequence of a property of the indicator function: for any two sets A, B, $[\mathbf{1}\{A\} - \mathbf{1}\{B\}]^2 = |\mathbf{1}\{A\} - \mathbf{1}\{B\}|$). Suppose that x_i^{d+1} is continuously distributed conditionally on $\mathcal{X}_i$. Then, except on a set of P-measure zero,

$$\begin{aligned}&\left|h_{\theta_1}(z_1, z_2) - h_{\theta_2}(z_1, z_2)\right| = \mathbf{1}\left\{y_1 \neq y_2\right\} \cdot \\ &\left|\mathbf{1}\{x_2^{d+1} > x_1^{d+1} + \theta_1'\left(\mathcal{X}_1 - \mathcal{X}_2\right)\} - \mathbf{1}\{x_2^{d+1} > x_1^{d+1} + \theta_2'\left(\mathcal{X}_1 - \mathcal{X}_2\right)\}\right|\end{aligned} \tag{16}$$

Suppose, further, that the density $\phi_{x^{d+1}|\mathcal{X}}$ is uniformly bounded and that components of $\mathcal{X}$ are P-integrable. Then

$$P^2\left[(h_{\theta_1}-h_{\theta_2})^2\right] \leq 2\left\|\theta_1-\theta_2\right\| P\left\|\mathcal{X}\right\| \sup \phi_{x^{d+1}|\mathcal{X}}. \tag{17}$$

This is inequality (13), because in the case of $m=2$, $P^{m-2}h_\theta = h_\theta$. The same inequality can be obtained without difficulty for all the other rank correlation estimators using similar considerations.

To prove the reverse inequality (15) for MRC, assume that x^{d+1} has a continuous density $\phi_{x^{d+1}|Y,\mathcal{X}}$ conditionally on both Y and $\mathcal{X}$. Then

$$P^2\left[(h_{\theta_1}-h_{\theta_2})^2\right] \geq \left\|\theta_1-\theta_2\right\| \cdot$$
$$\int \mathbf{1}\left\{Y_1 \neq Y_2\right\} \left|(\mathcal{X}_1-\mathcal{X}_2)'\mathbf{n}_{\theta_1-\theta_2}\right| \mu dP(Y_2,\mathcal{X}_2)dP(Y_1,X_1).$$

In this formula $\mathbf{n}_{\theta_1-\theta_2}$ is the unit vector in the direction of $\theta_1-\theta_2$, $P(Y_2,\mathcal{X}_2)$ and $P(Y_1,X_1)$ are marginal c.d.f.s of, respectively, $(Y,\mathcal{X})$ and (Y,X), and

$$\mu(X_1,Y_2,\mathcal{X}_2) = \min_{u\in\delta\|\mathcal{X}_1-\mathcal{X}_2\|\cdot[-1,1]} \phi_{x^{d+1}|Y_2,\mathcal{X}_2}\left(x_1^{d+1}+u\right),$$

where $\delta>0$ is so large that the compact Θ lies in the ball of radius δ. If $\phi_{x^{d+1}|Y,\mathcal{X}}$ is everywhere positive (so that $\mu>0$), and the set

$$\{Y_1 \neq Y_2,\ (\mathcal{X}_1-\mathcal{X}_2)'\mathbf{n}_{\theta_1-\theta_2} \neq 0\}$$

has a positive $P-$measure, then the reverse of (17) holds. For $m=2$, this is the same as inequality (15).

Inequality (15) can be verified by similar methods for the other rank estimators that maximize a U-processes of the second order, and for MR with the sample rank function S (which effectively maximizes a U-process of the third order), but *not* for the pairwise-difference estimators like PDR3 or PDR4. For the latter, (14) holds instead (for small differences $\theta_1-\theta_2$). This property lies at the origin of the improvement in accuracy of inference associated with the pairwise-difference rank estimators; see the next subsection.

In the bootstrap problem, we also need to account for the unconditional statistical dependence between the bootstrap draws.

Assumption 7 *There exist $\delta, C>0$ such that for all θ_1, θ_2 in the δ-neighborhood of θ_0,*

$$\left(P^{m-2}h_{\theta_1}^{[m-2]} - P^{m-2}h_{\theta_2}^{[m-2]}\right)^2 \leq C\left\|\theta_1-\theta_2\right\|.$$

Again, this condition is immediately true for MRC. For the other rank estimators it can be verified in the same way as Assumption 6, under the same sufficient conditions.

For estimators satisfying Assumption 6 (and, for the bootstrap, Assumption 7), the following bound holds.

Theorem 4 *Let Assumptions 1-3 and 6 hold. Assume that* $P\left|\sup_{\theta\in\Theta} r_{n,\theta}\right| = O\left(n^{-3/2}\right)$, $PM^2 < \infty$, $P\left\|\partial^2\tau_{\theta_0}\right\|^2 < \infty$, $P\left\|\partial\tau_{\theta_0}\right\|^4 < \infty$, *and, for a* $p \geq 6$, $P^m H^p < \infty$. *Then*

$$\sup_{A\in\mathcal{A}}\left|\int_A dF_{n^{1/2}(\theta_n-\theta_0)} - \int_A d\Phi_\Gamma\right| = O\left(\left(n^{-1/6}\left(\log n\right)^{2/3}\right)^{\frac{1}{1+2/3p}}\right). \tag{18}$$

If, additionally, $P\left|\sup_{\theta\in\Theta}\hat{r}_{n,\theta}\right| = O\left(n^{-3/2}\right)$, $P^m H^p_{\omega_m} < \infty$ *for each* ω_m, *and Assumption 7 holds, then*

$$\sup_{A\in\mathcal{A}}\left|\int_A d\hat{F}_{n^{1/2}\left(\hat{\theta}_n-\theta_n\right)} - \int_A d\Phi_\Gamma\right| = O_p\left(\left(n^{-1/6}\left(\log n\right)^{2/3}\right)^{\frac{1}{1+2/3p}}\right). \tag{19}$$

The error with which the bootstrap quantiles of $\hat{\theta}_n$ approximate the finite-sample quantiles of θ_n can be found from (18) and (19) by the triangle inequality. In the case of MRC ($P^m H^p$, $P^m H^p_{\omega_m} < \infty$ for all positive p), we have

$$\sup_{A\in\mathcal{A}}\left|\int_A dF_{n^{1/2}(\theta_n-\theta_0)} - \int_A d\hat{F}_{n^{1/2}\left(\hat{\theta}_n-\theta_n\right)}\right| = O_p\left(n^{-1/6+\varepsilon}\right),$$

where $\varepsilon > 0$ can be taken arbitrarily small.

2.3 Rates of Convergence: Pairwise-Difference Rank Estimators

The bound obtained in the previous subsection converges to zero slowly. The rate of convergence improves substantially, however, if the quantity on the left hand side of (13) has stronger continuity properties. Namely, let the following assumption hold.

Assumption 8 *For* $s = 2$ *or* 3[6], *function* $f_\theta = P^{m-s}h_\theta$ *is three times continuously differentiable in a* δ_0*-neighborhood of* θ_0. *There exists a function*

[6] If this condition is satisfied with $s > 3$, then it is also satisfied with $s = 3$, which is sufficient for our analysis.

$L(z_1,...,z_s)$, satisfying the condition $P^s L^4 < \infty$, such that

$$\|\partial f_{\theta_0}\|, \left\|\partial^2 f_{\theta_0}\right\|, \left\|\partial^3 f_{\theta_0}\right\| \leq L$$

and, for all θ_1, θ_2 in the δ_0-neighborhood of θ_0,

$$\left\|\partial^3 f_{\theta_1} - \partial^3 f_{\theta_2}\right\| \leq L \|\theta_1 - \theta_2\|.$$

(Here $\partial^k f_\theta$, $k \geq 3$, is the array of all partial derivatives of f of order k at θ, and $\left\|\partial^k f_\theta\right\|$ is the maximum in the absolute value over all elements of the array.)

It is clear that Assumption 8 cannot hold for MRC or for the variant of MR with the objective function given by a U-process of order 2, for which inequality (15) is true. Nonetheless, it can be satisfied by certain rank estimators maximizing a U-process of order 3 or higher. Historically, the first such example is Han's estimator of the parameter of the Box-Cox transformation function (Han (1987b)). Another example is a modification of Han's estimator proposed in Asparouhova et al. (2002). Abrevaya's PDR3 and PDR4 estimators also satisfy Assumption 8. Below we will concentrate on Abrevaya's estimators as they have broader applicability than the former two estimators.

For instance, consider the objective function of the PDR3 estimator. The symmetrized version of the kernel of the U-process is:

$$\begin{aligned} &h_\theta(z_1, z_2, z_3) = \\ &(\mathbf{1}\{y_1 > y_2\} - \mathbf{1}\{y_2 > y_3\}) \left(\mathbf{1}\{(x_1 - x_2)'\beta > (x_2 - x_3)'\beta\}\right) \\ &+ (\mathbf{1}\{y_2 > y_3\} - \mathbf{1}\{y_3 > y_1\}) \left(\mathbf{1}\{(x_2 - x_3)'\beta > (x_3 - x_1)'\beta\}\right) \\ &+ (\mathbf{1}\{y_3 > y_1\} - \mathbf{1}\{y_1 > y_2\}) \left(\mathbf{1}\{(x_3 - x_1)'\beta > (x_1 - x_2)'\beta\}\right) \end{aligned} \tag{20}$$

where the scale of β is normalized by setting $\beta = (\theta, 1)$. We will now check Assumption 8 with $s = 2$. To compute the value of function $P^{m-2} h_\theta$ one should integrate out the pair (x_3, y_3) in every term in the above expression. Once the $d+1$-th component of the vector of regressors is integrated out, the first term becomes

$$(\mathbf{1}\{y_1 > y_2\} - \mathbf{1}\{y_2 > y_3\}) \int_{(2x_2 - x_1)'\beta - \mathcal{X}_3'\theta}^{+\infty} \phi_{x^{d+1}|\mathcal{X}_3, y_3}(x)\, dx.$$

The derivative of this expression with respect to θ is

$$\begin{array}{l}(\mathbf{1}\{y_1 > y_2\} - \mathbf{1}\{y_2 > y_3\})(\mathcal{X}_1 + \mathcal{X}_3 - 2\mathcal{X}_2) \\ \cdot\phi_{x^{d+1}|\mathcal{X}_3,y_3}\left((2\mathcal{X}_2 - \mathcal{X}_1 - \mathcal{X}_3)'\theta + 2x_2^{d+1} - x_1^{d+1}\right).\end{array}$$

Similar expressions can be obtained for the other two terms in (20). The following conditions are sufficient for Assumption 8: $\phi_{x^{d+1}|\mathcal{X},Y}$ is three times differentiable in x^{d+1} for almost all $\mathcal{X}$ and Y and is uniformly bounded together with its derivatives of orders up to 3, and $P\|\mathcal{X}\|^{12} < \infty$. By a similar derivation, the PDR4 estimator (as well as the estimators of Han and Asparouhova et al.) satisfies Assumption 8, with $s = 3$, under the same sufficient conditions.

Not every rank estimator whose criterion function is a U-process of order three satisfies Assumption 8. Consider the MR estimator with the sample rank function S. After symmetrization,

$$\begin{aligned} h_\theta(z_1, z_2, z_3) &= \\ \mathbf{1}\{y_1 > y_3\}\mathbf{1}\{x_1'\beta &> x_2'\beta\} \\ +\mathbf{1}\{y_3 > y_2\}\mathbf{1}\{x_3'\beta &> x_1'\beta\} \\ +\mathbf{1}\{y_2 > y_1\}\mathbf{1}\{x_2'\beta &> x_3'\beta\}. \end{aligned}$$

The value of $P^{m-2}h_\theta$ is obtained by integrating out (x_3, y_3). After that, the first two terms will become differentiable in θ, while the last term will still contain the indicator function $\mathbf{1}\{x_2'\beta > x_1'\beta\}$. It is clear that under general conditions, inequality (15) will hold, which is incompatible with Assumption 8.

When Assumption 8 is satisfied, the components of θ_n that it controls decrease rapidly with n. The following condition is imposed to ensure a similar asymptotic behavior of the higher-order terms.

Assumption 9 *Either $m = s$ or there exist $\delta, C > 0$ such that for all θ_1, θ_2 in the δ-neighborhood of θ_0,*

$$P^{s+1}\left[\left(P^{m-(s+1)}h_{\theta_1} - P^{m-(s+1)}h_{\theta_2}\right)^2\right] \le C\|\theta_1 - \theta_2\|.$$

Similarly to the previous cases, an extra condition is needed in the bootstrap problem.

Assumption 10 *(a) Assumption 8 is satisfied with a function L such that, for every ω_s,*

$$P^s L_{\omega_s}^4 < \infty.$$

(b) If $s = 2$ or 3, and $m > s$, then there exist $\delta, C > 0$ such that for all θ_1, θ_2 in the δ-neighborhood of θ_0,

$$P^{s-1}\left[\left(P^{m-(s+1)}h_{\theta_1}^{[m-2]} - P^{m-(s+1)}h_{\theta_2}^{[m-2]}\right)^2\right] \leq C\left\|\theta_1 - \theta_2\right\|.$$

If $s = 3$ and $m > 3$, then, additionally,

$$\left(P^{m-4}h_{\theta_1}^{[m-4]} - P^{m-4}h_{\theta_2}^{[m-4]}\right)^2 \leq C\left\|\theta_1 - \theta_2\right\|,$$

where

$$\begin{aligned} & h_\theta^{[m-4]}\left(z_1, ..., z_{m-4}\right) \\ = & \int h_\theta\left(z_1, ...z_{m-4}, Z_{m-1}, Z_{m-1}, Z_m, Z_m\right) dP\left(Z_{m-1}\right) dP\left(Z_m\right). \end{aligned}$$

For PDR3 and PDR4 (and Han's and Asparouhova et al. estimators) these conditions can be checked, for, respectively, $s = 2$ and $s = 3$, by the same methods that were used to obtain (17). Moreover, in Assumption 9, generally the reverse inequality is also true, which can be verified by an argument similar to the proof of inequality (15) for MRC.

The next theorem gives the rates of convergence for rank estimators satisfying Assumptions 8 and 9. For brevity, only the case of uniformly bounded functions h is considered.

Theorem 5 *Suppose that Assumptions 1-3, 8 and 9 hold, $\sup_{Z,\theta}\left|r_{n,\theta}\right| = O\left(n^{-2}\right)$, and H is a constant. If Assumptions 8, 9 are satisfied with $s = 2$, let $\varepsilon > 0$ be arbitrarily small, and if they are satisfied with $s = 3$, let ε be zero. Then*

$$\sup_{A\in\mathcal{A}}\left|\int_A dF_{n^{1/2}(\theta_n-\theta_0)} - \int_A d\Phi_\Gamma\right| = O\left(n^{-1/2+\varepsilon}\right). \tag{21}$$

If also $\sup_{Z,\theta}\left|\hat{r}_{n,\theta}\right| = O\left(n^{-2}\right)$, Assumptions 5 (a) and 10 hold, then

$$\sup_{A\in\mathcal{A}}\left|\int_A d\hat{F}_{n^{1/2}\left(\hat{\theta}_n-\theta_n\right)} - \int_A d\Phi_\Gamma\right| = O_p\left(n^{-1/2+\varepsilon}\right). \tag{22}$$

3 Numerical Evidence

In this section we investigate the bootstrap properties of rank estimators in finite samples. We will consider two estimators: MRC and PDR4. As explained in the previous sections, MRC has a wider scope of applications (in particular, it can be applied to limited dependent variable models) and is cheaper to compute. However, the theory of Chapter 2 suggests that inference with MRC may be inaccurate in small samples. PDR4, on the other hand, needs substantial computational capacity (the fastest available algorithm for computing its objective function requires $O\left(n^2 \log n\right)$ operations and $O\left(n^2\right)$ memory cells). However, within the scope of its application, PDR4 can serve as a good complement to MRC in small samples, where it achieves higher precision of inference.

3.1 Monte-Carlo Experiments

In the Monte-Carlo experiments, MRC is applied to the binary choice model:

$$Y_i = \mathbf{1}\left\{X_i^{(1)} + X_i^{(2)} + \sigma\varepsilon_i > 0\right\}.$$

(In this case MRC and MR are numerically equivalent, so the evidence presented below illustrates the properties of MR as well.) Three distributions for the first regressor are considered: the standard normal (a continuous case), binomial (a discrete case) and the Student distribution with 1.5 degrees of freedom. In the latter case, the first moment of $X^{(1)}$ is finite, but its second moment is infinite. This is a situation where the nonparametric method for computing the asymptotic variance of the estimator will be rather difficult to apply. In particular, the moment conditions of Theorem 4 in Sherman (1993), under which the method is known to be consistent, are violated. Also, the rule for choosing the bandwidths (proportionally to the sample standard deviation of the estimated index $X'\hat{\beta}$) suggested by Cavanagh and Sherman (1998) can result in arbitrarily large bandwidths and is not practical. The second regressor, $X^{(2)}$, is distributed as $N\left(0,1\right)$ independently of $X^{(1)}$. It plays the role of a continuously distributed regressor needed for point identification of β. The error term, ε, is also distributed as $N(0,1)$ independently of both regressors. The scaling parameter σ, therefore, determines the noise-to-signal ratio in the dataset. We consider two cases, $\sigma = 1$ and $\sigma = 0.1$.

PDR4 is applied to the linear model:

$$Y_i = X_i^{(1)} + X_i^{(2)} + \varepsilon_i.$$

The regressor $X^{(1)}$ can have the standard normal or the Student(1.5) distribution. Regressor $X^{(2)}$ is distributed as a standard normal random variable independently of $X^{(1)}$. The error term is independent of both regressors and is distributed as either a standard normal or a standard Cauchy random variable. The latter case serves to demonstrate the robustness properties of PDR4 with respect to large errors. Note that for the Cauchy distributed errors, $P\,|\varepsilon| = +\infty$, so that the OLS or nonparametric minimum-square-distance methods are not consistent in this case.

In rank regressions, identification is achieved by imposing a scale normalization on the vector of estimated coefficients. Here we set the coefficient at the second regressor to be 1. The estimated model is then

$$y_i = f\left(\theta X_i^{(1)} + X_i^{(2)} + \varepsilon_i\right)$$

where $f(x) = 1\{x > 0\}$ in the binary choice model, and $f(x) = x$ in the linear model (function f does not have to be known for implementation of MRC or PDR4) and ε_i is the error term. The value of θ is found by maximizing the corresponding criterion function. The MRC objective function is rather nonsmooth for our sample sizes, and its maximization is more difficult than that of the PDR4 objective function. We used the Nelder-Mead simplex maximization algorithm with parameters adjusted in trial runs of the program. For PDR4 estimator the standard maximization Matlab routine fminsearch with default settings was enough. Both algorithms are iterative procedures requiring a starting approximation of the solution. For the population problem, we took the true value $\theta_0 = 1$. This option, of course, is not available in real data applications, where a grid of initial values should be considered. In the bootstrap we used both θ_0 and θ_n.

There are several asymptotically equivalent methods of computing the bootstrap critical values for the test statistic $n^{1/2}(\theta_n - \theta_0)$ that do not need an explicit estimator of the asymptotic variance. In the percentile method the quantiles of the test statistic are approximated by the conditional quantiles of the bootstrapped recentered statistics $n^{1/2}\left(\hat{\theta}_n - \theta_n\right)$. In our experiments with MRC, however, recentering of the bootstrapped estimator at θ_n led to relatively inaccurate results. One alternative, based on the symmetry of the normal distribution, is the other percentile method, see Hall (1992), in which the quantiles of the test statistics are approximated by the quantiles of the statistic $n^{1/2}\left(\theta_n - \hat{\theta}_n\right)$. This procedure effectively eliminates recentering since θ_n cancels out in the resulting confidence intervals and criteria for hypothesis testing. This method was used to compute one-sided (left-tailed and right-tailed) and double-sided (equal-tailed) critical values. The rejection probabilities for the corresponding tests were similar, and, for the sake of brevity, we report them only for the double-sided case[7]. There are other procedures that do not require recentering at θ_n. One can approximate the c.d.f. of the test statistic by the c.d.f. of the demeaned bootstrapped statistic, $n^{1/2}\left(\hat{\theta}_n - \hat{P}\left[\hat{\theta}_n\right]\right)$, or by the c.d.f. of the normal distribution with zero mean and variance estimated by the conditional variance of $n^{1/2}\hat{\theta}_n$. These two methods will be more convenient than the other percentile method for inference about multidimensional θ. The results for both are similar to the case of the other percentile method and are omitted.

MRC was computed for sample sizes $n = 200, 500$, and 1000 (see Table 1). The coverage probabilities are reasonably accurate except in the case with $n = 200$ and $\sigma = 0.1$ where the bootstrap fails dramatically for all three distributions of $X^{(1)}$. This should serve as a caution against using the bootstrap when the signal-to-noise ratio is high and the sample size is moderate. In this case the simulated distribution of MRC appears to have a mass point at zero. The bootstrap gives a distribution with a much higher concentration of mass at zero, and so underestimates both the quantiles and the variance of the estimator. The phenomenon has to be taken into account when MRC is used together with a specification search: the bootstrap may reject models with low noise more often than it should.

In the case of PDR4, the percentile method (involving recentering) and the other percentile method gave close values of rejection probabilities. For brevity we report only the values obtained for the equal-tailed tests by the other percentile method, for sample sizes $n = 50$, 100, and 200 (Table 2). It can be seen that bootstrap performs well even when the sample includes only 50 observations.

[7] For a description of how rejection probabilities are computed, see Hall and Horowitz (1996).

Table 1. Rejection probabilities for equal-tailed t-test - MRC

	nominal level 5%			nominal level 10%		
$n =$	200	500	1000	200	500	1000
Normal $X^{(1)}$						
$\sigma = 1$	3.0	3.2	5.1	7.7	8.5	9.4
$\sigma = 0.1$	25.0	5.9	3.6	31.2	10.6	8.1
Binary $X^{(1)}$						
$\sigma = 1$	3.6	4.2	4.2	7.7	9.6	8.3
$\sigma = 0.1$	31.9	5.5	2.8	35.0	9.2	5.9
Student (1.5) $X^{(1)}$						
$\sigma = 1$	2.7	2.5	4.3	5.3	7.3	8.9
$\sigma = 0.1$	33.8	6.8	3.5	38.6	10.1	7.2

Table 2. Rejection probabilities for equal-tailed t-test - PDR4

	nominal level 5%			nominal level 10%		
$n =$	50	100	200	50	100	200
Normal $X^{(1)}, \varepsilon$	5.9	4.8	5.8	12.8	10.1	10.3
Student (1.5) $X^{(1)}, \varepsilon$	6.4	4.9	4.6	12.5	10.8	10.7
Normal $X^{(1)}$, Cauchy ε	6.1	5.0	6.2	13.9	9.6	9.5

3.2 Empirical Example

The main purpose of this subsection is to provide a practical sense for using the bootstrap for rank estimators in real-data applications. We continue the wage-equation example studied by Abrevaya (2003). The data set, constructed by Ruud (2000), is an extract from the March 1995 CPS, consisting of 1,289 observations. The dependent variable is an hourly wage (WAGE). The regressors are years of schooling (EDUC), years of potential work experience (EXPER) and its square (EXPSQ), a female indicator variable (FEMALE), a union indicator variable (UNION), and a nonwhite indicator variable (RACE equal to 0 if white, 1 if not). The model is specified as the transformation model with an unknown link function f:

$$\begin{aligned} f(WAGE) &= \beta_1 EDUC + \beta_2 EXPER + \beta_3 EXPSQ \\ &\quad + \beta_4 FEMALE + \beta_5 UNION + \beta_6 RACE + \varepsilon \end{aligned}$$

(the traditional choice of f in such models is the logarithmic function). The identification assumption is that f is a strictly increasing function and ε is an i.i.d. error term distributed independently of the regressors.

The coefficients are estimated by MRC and PDR4, with a scale normalization $\beta_1 = 1$. Abrevaya (2003) computed the estimates of the coefficients. He also applied the nonparametric method to estimate their standard errors. As noted in Introduction, the nonparametric method is computationally intensive, requiring $O(n^2)$ computations in the case of MRC and $O(n^4)$ computations in the case of PDR4. Also, it involves $3+d$ one-dimensional nonparametric regressions (where $d+1$ is the dimension of the vector of covariates), so that in the example considered here one has to run eight kernel regressions[8]. As the implementation of each of them requires a choice of a bandwidth and some other details (such as the form of the kernel and the trimming parameters in the denominator of the Nadaraya-Watson conditional expectation estimator), the method contains an element of subjectivity.

Here we provide alternative estimates of the standard errors obtained by the bootstrap. The computational burden of the bootstrap is of order $O(Bn\log n)$ for MRC and $O(Bn^2\log n)$ for PDR4, where B is the number of bootstrap iterations. We used $B = 1000$, although the estimates of the standard errors did not change much after $B = 200$ iterations already. For this sample size, the computation of the MRC objective function is much faster than that of the PDR4 objective function, but the associated maximization problem for the former is more difficult to solve numerically. In the case of MRC we used the Nelder-Mead algorithm of optimization with five different combinations of parameters and starting values (chosen in trial runs) in each bootstrap iteration. For PDR4 we used the standard Matlab maximization routine fminsearch with default settings and one starting vector of parameters, the estimated vector. The computational times for one thousand bootstrap iterations were 34 minutes for MRC and 6.5 hours for

[8] For the PDR4 estimator, estimation of the asymptotic variance using formulas (A.13)-(A.14) in Abrevaya (2003) involves $4 + \frac{d(d+1)}{2}$ one-dimensional nonparametric regressions. However, computing the nonparametric estimate of the conditional variance matrix in these formulas can be avoided, and so the number can be reduced to $3 + d$.

PDR4, on an AMD Opteron 2.8 GHz processor[9]. The memory usage was 62 megabytes for MRC and about 400 megabytes for PDR4.

Table 3, reports the values of the estimated coefficients and the standard errors. It can be seen from this example that the bootstrap may give the standard errors that are substantially different from the ones obtained by the nonparametric method.

Table 3. Wage-Equation Estimation Results

	MRC			PDR4		
	coef.	std. error		coef.	std. error	
		nonpar.	boots.		nonpar.	boots.
$EDUC$	1.0000	-	-	1.000	-	-
$EXPER$	.3590	.0559	.0502	.4068	.0487	.0432
$EXPSQ$	-.5965	.1251	.1123	-.6741	.1140	.0977
$FEMALE$	-2.2105	.3187	.2744	-2.3252	.3102	.2747
$RACE$	-.9851	.4537	.2937	-1.2828	.4076	.3492
$UNION$	1.5178	.4482	.3022	1.8922	.4103	.3166

4 Conclusion

This paper provides the bootstrap theory and extends the asymptotic theory of rank estimators, a class of methods that can be applied to popular semiparametric single-index models or used for robust estimation of parametric models. Under general regularity conditions, rank estimators were previously known to have an asymptotic normal distribution. Here we proved that the parameters of that distribution can be estimated by the nonparametric bootstrap. With the bootstrap, estimation and inference in rank regressions is entirely free of any tuning parameters, a property not enjoyed by other available semiparametric techniques. We have investigated the accuracy of such inference and provided bounds on the associated error. In the case of MRC or MR, the bound is a function of the sample size of order close to $n^{-1/6}$. Pairwise-difference rank estimators, however, have a special structure due to which the bound is vanishing with the rate close to $n^{-1/2}$. Thus, pairwise-difference estimators provide a remarkable example of a robust semiparametric method whose first- and second-order asymptotic properties approach those of para-

[9] The objective functions of the two estimators were computed using the fast algorithm of Abrevaya (1999) for MRC and a sorting-based algorithm for PDR4, both programmed in C. The codes, compiled as Matlab functions, are available from the author upon request.

metric methods. We have illustrated our theoretical results in finite-sample Monte-Carlo experiments, and demonstrated their practical usefulness in an empirical example.

Acknowledgement

The author is grateful to Joel Horowitz and Elie Tamer for numerous helpful discussions of the material presented in the paper and encouragement; Jason Abrevaya, Federico Bugni, Shakheeb Khan, Rosa Matzkin, Robert Sherman, participants of the 2006 Annual Meeting of the Midwest Econometric Group, the 2007 Meeting on Semiparametric and Nonparametric Methods in Econometrics at Obervolfach (Germany), the 2007 North American Summer Meeting of the Econometric Society, the econometrics seminar at California Institute of Technology, and the economics seminars at the the Pennsylvania State University, the University of Texas at Austin, Yale University, and the Center for Economic and Financial Research at New Economic School (Moscow, Russia) for important suggestions and comments.

References

Abadie A., and Imbens, G. (2006). On the failure of the bootstrap for matching estimators, mimeo.

Abrevaya, J. (1999). Computation of the Maximum Rank Correlation Estimator. Economics Letters, 62, 279-285.

Abrevaya, J. (1999a). Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable. Journal of Econometrics, 93, 203-228.

Abrevaya, J. (1999b). Rank estimation of a transformation model with observed truncation. Econometric J., 2, 292-305.

Abrevaya, J. (2003). Pairwise-difference rank estimation of the transformation model. J. Bus. Econom. Statist. 21, no. 3, 437–447.

Ai, C.; Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica 71, no. 6, 1795–1843.

Arcones, M. A. (1995). The asymptotic accuracy of the bootstrap of U-quantiles. Ann. Statist. 23, no. 5, 1802–1822.

Arcones, M. A.; Chen, Z.; Giné, E. (1994). Estimators related to U-processes with applications to multivariate medians: asymptotic normality. Ann. Statist. 22, no. 3, 1460–1477.

Arcones, M. A.; Giné, E. (1992). On the bootstrap of U and V-statistics. Ann. Statist. 20, no. 2, 655–674.

Arcones, M. A.; Giné, E. (1993). Limit theorems for U-processes. Ann. Probab. 21, no. 3, 1494–1542.

Arcones, M. A.; Giné, E. (1994). U-processes indexed by Vapnik-Červonenkis classes of functions with applications to asymptotics and bootstrap of U-statistics with estimated parameters. Stochastic Process. Appl. 52, no. 1, 17–38.

Asparouhova, E.; Golanski, R.; Kasprzyk, K.; Sherman, R. P.; Asparouhov, T. (2002). Rank estimators for a transformation model. Econometric Theory 18, no. 5, 1099–1120.

Bhattacharya R., and Ranga Rao R. (1976). Normal Approximation and Asymptotic Expansions, Wiley, New York.

Bickel, P.J., and Freedman, D.A. (1981). Some Asymptotic Theory for the Bootstrap. The Annals of Statistics, 9, 1196-1217.

Bolthausen, E.; Götze, F. (1993). The rate of convergence for multivariate sampling statistics. Ann. Statist. 21, no. 4, 1692–1710.

Cavanagh, C., and Sherman, R. P. (1998). Rank Estimators for Monotonic Index Models. Journal of Econometrics, 84, 351-381.

Chen, S. (2002). Rank estimation of transformation models. Econometrica 70, no. 4, 1683–1697.

De Angelis, D., Hall, P., and Young G.A. (1993). Analytical and Bootstrap Approximations to Estimator Distributions in L^1 Regressions. Journal of the American Statistical Association, 88, 1310-1316.

De la Peña, V. H. (1992). Decoupling and Khintchine's inequalities for U-statistics. Ann. Probab. 20, no. 4, 1877–1892.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. Ann. Statist. 7, no. 1, 1–26.

Giné, E.; Mason, D. M. (2007). On local U–statistic processes and the estimation of densities of functions of several sample variables. Annals of Statistics, to appear.

Giné E. and Zinn J. (1990). Bootstrapping general empirical measures. Annals of Probability 18, 851-869.

Giné E. and Zinn J. (1992). On Hoffmann-Jørgensen's inequality for U-processes. In Probability in Banach spaces 8, Birkhauser Progress in Probability Series, Vol. 30, pp. 80-91, (R. Dudley, J. Kuelbs, M. Hahn eds.), Boston.

Hall, P. (1992). The bootstrap and Edgeworth expansion. Springer Series in Statistics. Springer-Verlag, New York.

Hall P., and Horowitz J. (1996). Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators, Econometrica, 64, 891-916.

Han, A. K. (1987). Non-Parametric Analysis of a General Regression Model. The Maximum Rank Correlation Estimator. Journal of Econometrics, 35, 303-316.

Han, A. K. (1987b). A nonparametric analysis of transformations. J. Econometrics 35, no. 2-3, 191–209.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. J. Econometrics 58, no. 1-2, 71–120.

Khan S., and Tamer E. (2007). Partial Rank Estimation of Duration Models with General Forms of Censoring. Journal of Econometrics 136, 251-280.

Klein R.W., and Spady R.H. (1993). An Efficient Semiparametric Estimator for Binary Response Models. Econometrica, 61, 387-421.

Lee, M. (1999) A root-N consistent semiparametric estimator for related-effect binary response panel data. Econometrica, 67, 427–433.

Nishiyama, Y.; Robinson, P. M. (2005). The bootstrap and the Edgeworth correction for semiparametric averaged derivatives. Econometrica 73, no. 3, 903–948.

Nolan, D., and Pollard, D. (1987). U-Processes: Rates of Convergence. The Annals of Statistics, 15, 780–799.

Pakes, A., and Pollard, D. (1989). Simulation and the Asymptotics of Optimization Estimators. Econometrica, 57, 1027–1057.

Pollard, D. (1985). New ways to prove central limit theorems. Econometric Theory, 1, 295–313.

Pollard, D. (1989). Asymptotics via Empirical Processes (with Discussion). Statistical Science, 4, 341–366.

Powell J. L., Stock J. H., and Stoker T. M. (1989). Semiparametric Estimation of Index Coefficients. Econometrica, 57, 1403-1430.

Ruud, P. A. (2000), An Introduction to Classical Econometric Theory, Oxford, U.K.: Oxford University Press.

Serfling, R.J. (1980). Approximation Theorems of Mathematical Statistics. New York: Wiley.

Sherman, R.P. (1993). The Limiting Distribution of the Maximum Rank Correlation Estimator. Econometrica, 61, 123-137.

Sherman, R.P. (1994). Maximal Inequalities for Degenerate U-Processes with Applications to Optimization Estimators. Annals of Statistics, 22, 439-459.

Van der Vaart, A.W., and Wellner, J.A. (1996). Weak Convergence and Empirical Processes. New York: Springer-Verlag.

5 Appendix

Appendix contains the proofs of Theorems 1-5. It consists of four parts. The structure of the proofs and their main ingredients are discussed in Section 5.1, and the actual proofs are in Section 5.2. Section 5.3 briefly reviews the known results from the empirical process theory for U-processes, and provides the necessary extensions. Section 5.4 contains an auxiliary lemma related to the Berry-Esséen bound for M-estimators with a criterion function in the form of a smooth U-process. Short versions of the proofs are presented; detailed proofs are available from the author upon request.

5.1 Main Tools of Proof

This subsection describes the main ideas underlying the proofs of Theorems 1-5. The essence of the analysis is to separate a smooth and a non-smooth components of the objective function. The estimator θ_n is approximated by a maximizer of the smooth component whose properties can be studied using the Taylor expansion and the Berry-Esséen bounds. Then the empirical process theory for U-processes is used to show that the contribution of the nonsmooth remainder in the objective function to the distribution of θ_n is small.

To simplify notation, we assume, without loss of generality, that $\theta_0 = 0$, and that the function h_0 is identically zero.

5.1.1 Approximation

Consider an estimator, θ_n, that solves the problem

$$G_{n,\theta_n} \geq \sup_{\theta \in \Theta} \left[G_{n,\theta} - r_{n,\theta} \right],$$

and assume that the objective function $G_{n,\theta}$ admits the representation

$$G_{n,\theta} = G^0_{n,\theta} + \zeta_{n,\theta}, \tag{23}$$

where $\theta \in \Theta \subset \mathbb{R}^d$, $G^0_{n,\theta}$ is a smooth random function of θ, and $\zeta_{n,\theta}$ is a remainder. An approximation to θ_n, denoted by η_n, solves the problem

$$\eta_n \in \arg\max_{\theta \in \Theta} G^0_{n,\theta}. \tag{24}$$

If the remainder terms $\zeta_{n,\theta}$, $r_{n,\theta}$ are small in an appropriate sense, then the difference $n^{1/2}(\theta_n - \eta_n)$ will also be small. The following theorems formalize this idea.

The first theorem is useful for establishing the asymptotic normality of θ_n (part (a)), and estimating its variance (part (b)). Here it is enough to consider the representation (23) with

$$G^0_{n,\theta} \equiv \theta' W_n - \frac{1}{2}\theta' A \theta$$

where W_n is a $d \times 1$ random vector, not depending on θ, and A is a matrix of constants. Then $\eta_n = A^{-1} W_n$, as long as the vector on the right-hand side is an element of Θ. The first part of the theorem is a variant of Pollard's (1985) asymptotic normality theorem (see also Sherman (1993) and Arcones, Chen, Giné (1994)), and the second part is a simple extension.

Theorem 6 *Assume that* 0 *is an interior point of* Θ, *and* A *is a symmetric, positive definite, constant matrix. (a) If* $\theta_n \rightarrow^p 0$, $W_n = O_p\left(n^{-1/2}\right)$, *and for every sequence of numbers* $\delta_n \rightarrow +0$,

$$\sup_{\|\theta\| \le \delta_n} \frac{\left|\zeta_{n,\theta}\right| + \left|r_{n,\theta}\right|}{n^{-1} + \|\theta\|^2} \rightarrow^p 0, \tag{25}$$

then

$$n^{1/2}\left(\theta_n - A^{-1}W_n\right) \rightarrow^p 0.$$

(b) If, additionally, Θ *is a bounded set,* $P\|W_n\|^2 = O\left(n^{-1}\right)$, *and for every* $\varepsilon > 0$, *and every sequence of numbers* $\delta_n \rightarrow +0$,

$$P\left\{\|\theta_n\| > \varepsilon\right\} = o\left(n^{-1}\right),$$

$$P\left\|W_n^2\right\| 1\left\{\|W_n\| > \varepsilon\right\} = o\left(n^{-1}\right),$$

and

$$P\left\{\sup_{\|\theta\| \le \delta_n} \frac{\left|\zeta_{n,\theta}\right| + \left|r_{n,\theta}\right|}{n^{-1} + \|\theta\|^2} > \varepsilon\right\} = o\left(n^{-1}\right), \tag{26}$$

then

$$P\left\|n^{1/2}\left(\theta_n - A^{-1}W_n\right)\right\|^2 \rightarrow 0. \tag{27}$$

Proof. (Sketch.) Denote $t_n = n^{1/2}\left(\theta_n - A^{-1}W_n\right)$. When $A^{-1}W_n \in \Theta$, by the defining property of θ_n and (23),

$$t_n' A t_n \le 2n\left[\zeta_{n,A^{-1}W_n + n^{-1/2}t_n} - \zeta_{n,A^{-1}W_n} - r_{n,A^{-1}W_n}\right]. \tag{28}$$

Fix $\varepsilon > 0$, and let $E_{\varepsilon,n}$ be the event that

$$\|t_n\|^2 \le \varepsilon\left(1 + \|t_n\|^2 + n\|W_n\|^2\right).$$

(a) By the assumptions in part (a) of the theorem and (28), $P\left(\overline{E_{\varepsilon,n}}\right) = o(1)$, which implies that $t_n = o_p(1)$.

(b) Assumptions of part (b) of the theorem and (28) imply that $P\left(\overline{E_{\varepsilon,n}}\right) = o\left(n^{-1}\right)$. Choose $\varepsilon < 1$. We have:

$$
\begin{aligned}
& P\left\|t_n\right\|^2 \leq P\left\|t_n\right\|^2 1_{E_{\varepsilon,n}} + P\left\|t_n\right\|^2 1_{\overline{E_{\varepsilon,n}}} \\
\leq \; & \frac{\varepsilon}{1-\varepsilon} O(1) + 2nP\left\|\theta_n\right\|^2 1_{\overline{E_{\varepsilon,n}}} + 2nP\left\|W_n\right\|^2 1_{\overline{E_{\varepsilon,n}}}
\end{aligned}
$$

Since $\theta_n \in \Theta$ is bounded, $nP\left\|\theta_n\right\|^2 1_{\overline{E_{\varepsilon,n}}} = O(1) \cdot nP\left\{\overline{E_{\varepsilon,n}}\right\} = o(1)$. For the last term,

$$
\begin{aligned}
nP\left\|W_n\right\|^2 1_{\overline{E_{\varepsilon,n}}} & \leq nP\left\|W_n\right\|^2 \left\{\left\|W_n\right\|^2 > 1\right\} + nP\left\{\overline{E_{\varepsilon,n}}\right\} \\
& = o(1).
\end{aligned}
$$

Therefore, $P\left\|t_n\right\|^2 = o(1)$. ■

To assess the accuracy of the normal approximation, one needs to investigate the nature of the difference between θ_n and η_n more closely.

Theorem 7 . *Suppose that equations (23) and (24) hold. Assume that there exists a sequence of numbers $a_n \geq 1$, and numbers $\lambda, \delta_0 > 0$ and $\alpha \in [0,2)$ such that the ball with center zero and radius δ_0 is in Θ, and*
(i) For any $\delta > 0$, $P\left\{\left\|\eta_n\right\| + \left\|\theta_n\right\| > \delta\right\} = O\left(a_n^{-1}\right)$.
(ii)

$$
\begin{aligned}
& P\left\{\begin{array}{c} \text{Matrix } \partial^2 G^0_{n,\theta} \text{ exists and is continuous, and} \\ -\partial^2 G^0_{n,\theta} \geq \lambda I \text{ for all } \|\theta\| \leq \delta_0 \end{array}\right\} \\
= \; & 1 - O\left(a_n^{-1}\right).
\end{aligned}
$$

(iii) For any $0 < \delta \leq \delta_0$,

$$
\begin{aligned}
& P\left\{\sup_{\|\theta\| \leq \delta} \frac{\zeta_{n,\eta_n+\theta} - \zeta_{n,\eta_n} + r_{n,\eta_n}}{n^{-1}a_n^{-2} + \delta\|\theta\|^2 + \left(n^{-1/2}a_n^{-1}\right)^{2-\alpha}\|\theta\|^\alpha} \leq \frac{1}{\delta_0}\right\} \\
= \; & 1 - O\left(a_n^{-1}\right).
\end{aligned}
$$

Then there exists a constant K such that

$$
P\left\{n^{1/2}\left\|\theta_n - \eta_n\right\| > K a_n^{-1}\right\} = O\left(a_n^{-1}\right).
$$

Proof. Let $\delta^* = \min\left\{\delta_0, \frac{\lambda\delta_0}{4}\right\}$. Let E_n be the union of event

$$\{\|\theta_n\|, \|\eta_n\| < \delta^*\},$$

the event in condition (ii), and the event in condition (iii) for $\delta = \delta^*$. Conditions (i)-(iii) imply that $P\left(\overline{E_n}\right) = O\left(a_n^{-1}\right)$. Define $t_n = n^{1/2} a_n (\theta_n - \eta_n)$. Since $\eta_n \in \Theta$, we have

$$G^0_{n,\theta_n} - G^0_{n,\eta_n} \geq -r_{n,\theta} + \zeta_{n,\eta_n} - \zeta_{n,\theta_n}.$$

When on E_n, η_n is an interior maximum and so the F.O.C., $\partial G^0_{n,\eta_n} = 0$, is satisfied. Use this to expand the left-hand side around η_n: for some $\left\|\tilde{\theta}_n\right\| \leq \delta^*$,

$$\begin{aligned} G^0_{n,\theta_n} - G^0_{n,\eta_n} &= \frac{1}{2} n^{-1} a_n^{-2} t_n' \partial^2 G^0_{n,\tilde{\theta}_n} t_n \\ &\leq -\frac{\lambda}{2} n^{-1} a_n^{-2} \|t_n\|^2. \end{aligned}$$

Therefore,

$$\begin{aligned} \|t_n\|^2 &\leq \frac{2}{\lambda} n a_n^2 \left[r_{n,\theta} + \zeta_{n,\theta_n} - \zeta_{n,\eta_n}\right] \\ &\leq \frac{2}{\lambda\delta_0}\left[1 + \delta^* \|t_n\|^2 + \|t_n\|^\alpha\right] \end{aligned}$$

or (recall that $\delta^* \leq \frac{\lambda\delta_0}{4}$)

$$\|t_n\|^2 \leq \frac{4}{\lambda\delta_0}\left(1 + \|t_n\|^\alpha\right)$$

Because $\alpha \in [0, 2)$, this implies that for some constant $K = K(\alpha, \delta_0, \lambda) > 0$,

$$\|t_n\| \leq K.$$

Taking into account the possibility of the event $\overline{E_n}$, we have

$$P\left\{n^{1/2}\|\theta_n - \eta_n\| > K a_n^{-1}\right\} = P\left(\overline{E_n}\right) = O\left(a_n^{-1}\right).$$

■

5.1.2 Hoeffding Decomposition and Its Bootstrap Version

When the estimator maximizes a U-process, representation (23) can be obtained by the so-called Hoeffding decomposition (or the U-decomposition). Let $h_\theta : \mathcal{Z}^m \to \mathbb{R}$ be a symmetric, P-measurable function. Denote by $\pi_{k.m} h_\theta$ the projection of h_θ onto the space of functions of k arguments that are *degenerate* with respect to the measure P, in the sense that their expectation relative to P over any one argument, holding the other arguments constant, is zero:

$$(\pi_{k,m} h_\theta)(z_1, ..., z_k) = (\delta_{z_1} - P) ... (\delta_{z_k} - P) P^{m-k} h_\theta$$

(where $\delta_{z_1} h_\theta = h_\theta(z_1, \cdot)$). Then

$$U_n^{(m)} h_\theta = P^m h_\theta + m P_n \pi_{1,m} h_\theta + \sum_{k=2}^{m} \binom{m}{k} U_n^{(k)} (\pi_{k,m} h)_\theta . \tag{29}$$

where P_n is the sample mean, i.e. the U-process of order 1 (see e.g. Arcones and Giné (1992) for the U-decomposition in this notation).

The importance of the Hoeffding decomposition is that it isolates terms of progressively higher order in $n^{-1/2}$. The first term is the expectation of h_θ and has the order $O(1)$. The second term is the sample mean of a random variable with zero mean; it has order $O_p\left(n^{-1/2}\right)$ by the Central Limit Theorem. The following terms are of order $O_p\left(n^{-k/2}\right)$. Representation (23) can be obtained if the first few terms in (29) are twice differentiable in θ, so that they admit a Taylor expansion with leading terms given by $G^0_{n,\theta}$, while the error term $\zeta_{n,\theta}$ will collect the remainder from the Taylor expansion and the higher-order U-processes in (29). Specific decompositions will be considered below.

A similar decomposition is also needed for the bootstrap problem. In the literature on the bootstrap of U-statistics, it is common to write the Hoeffding decomposition of the bootstrapped process, $\hat{U}_n^{(m)} h_\theta$, conditionally on the sample of data $\{Z_i\}_{i=1}^n$, i.e. relative to the empirical measure P_n in place of P. This approach makes the analysis of the higher-order processes no more difficult in the bootstrap problem than in the sample problem. It is inconvenient for M-estimators, however, because the leading terms of the U-decomposition relative to P_n may not have the smoothness properties of the leading terms in (29). For example, the first term will be:

$$P_n^m h_\theta \equiv \frac{1}{n^2} \sum_{i_1,...,i_m} h_\theta(Z_{i_1}, Z_{i_2}, ..., Z_{i_m})$$

which is not a differentiable function of θ for the rank estimators. Thus, the Taylor expansion arguments leading to representation (23) for the sample problem will not be directly applicable to the bootstrap problem.

Here we suggest a different approach. Write the Hoeffding decomposition of the bootstrapped process in terms of the same functions $\pi_{k.m}h_\theta$ (integrals of h_θ relative to P) that appear in (29):

$$\hat{U}_n^{(m)}h_\theta = P^m h_\theta + m\hat{P}_n\left(\pi_{1,m}h_\theta\right) + \sum_{k=2}^{m}\binom{m}{k}\hat{U}_n^{(k)}\left(\pi_{k.m}h_\theta\right). \tag{30}$$

(To obtain this formula, apply the summation operator $\hat{U}_n^{(m)}$ to formula (2.5) in Arcones and Giné (1992).) Now, the functional form, and therefore, smoothness properties with respect to θ, of the leading terms in $G_{n,\theta}$ and $\hat{G}_{n,\theta}$ are the same, and only the sample of data on which they are evaluated differ.

5.1.3 Bounds on the Higher-Order U-Processes

To apply the approximation theorems, we need to check their equicontinuity assumptions for the components of $\zeta_{n,\theta}$ and $\hat{\zeta}_{n,\theta}$ given by the higher-order U-processes in the Hoeffding decomposition. This is the most challenging part of the proof, which is mostly deployed in Section 5.3. Here we give only the final results relevant to our problem.

Given a function $h\left(z_1,...,z_m\right)$, define the function

$$h^{[m-2s]}\left(z_1,...,z_{m-2s}\right) =$$
$$\int h\left(z_1,...,z_{m-2s},Z_{m-s+1},Z_{m-s+1},...,Z_m,Z_m\right)dP\left(Z_{m-s+1}\right)...dP\left(Z_m\right).$$

For the sample problem, the following two bounds hold.

Lemma 8 *(a) Let $\mathcal{H}=\{h_\theta:\mathcal{Z}^m\to\mathbb{R}\}$, $m\geq 1$, be a class of P-degenerate symmetric functions, which is Euclidean for an envelope H satisfying $P^mH^{p\vee 2}<\infty$ for $p\geq 1$, and $\mathcal{H}_n$ be its subclasses. Then, as $n\to\infty$,*

$$n^{m/2}\left(P\sup_{h\in\mathcal{H}}\left|U_n^{(m)}h\right|^p\right)^{1/p} = O\left(1\right).$$

(b) If, additionally, $\sup_{h\in\mathcal{H}_n} P^m h^2 \to 0$*, then*

$$n^{m/2}\left(P\sup_{h\in\mathcal{H}_n}\left|U_n^{(m)}h\right|^p\right)^{1/p} = o\left(1\right).$$

(c) If, additionally to conditions in (a), $P^m H_{\omega_m}^{p\vee 2} < \infty$ *for each* ω_m*, then*

$$n^{m/2}\left(P\sup_{h\in\mathcal{H}}\left|\hat{U}_n^{(m)}h\right|^p\right)^{1/p} = O\left(1\right).$$

(d) If, additionally to conditions in (b) and (c), for each s*,* $1 \le s \le \frac{m}{2}$*,* $\sup_{h\in\mathcal{H}_n} P^{m-2s}\left(h^{[m-2s]}\right)^2 \to 0$*, then*

$$n^{m/2}\left(P\sup_{h\in\mathcal{H}_n}\left|\hat{U}_n^{(m)}h\right|^p\right)^{1/p} = o\left(1\right).$$

Lemma 9 *Let* $\mathcal{H} = \{h_\theta : \mathcal{Z}^m \to \mathbb{R}\}$*,* $m \ge 2$*, be a class of symmetric,* P*-degenerate functions, Euclidean for an envelope* H*. Assume that there exist constants* δ_0*,* $C > 0$ *such that for all* θ_1*,* θ_2 *in the* δ_0*-neighborhood of* 0*,*

$$P^m\left[\left(h_{\theta_1} - h_{\theta_2}\right)^2\right] \le C\left\|\theta_1 - \theta_2\right\|. \tag{31}$$

Then

$$P\left\{\sup_{\|\tilde{\theta}\|,\ \|\theta\|\le\delta_0/2}\frac{\left|U_n^{(m)}\left(h_{\tilde{\theta}+\theta} - h_{\tilde{\theta}}\right)\right|}{n^{-1}a_n^{-2} + \left(n^{-1/2}a_n^{-1}\right)^{3/2}\|\theta\|^{1/2}} > 1\right\} = O\left(a_n^{-1}\right), \tag{32}$$

with any $a_n \ge 1$ *satisfying*

$$a_n \le \left(n^{1/6}\left(\log n\right)^{-2/3}\right)^{1/(1+2/3p)},\ \text{ if } m = 2 \text{ and } P^m H^6 < \infty;$$

$$a_n \le n^{(m-1)/4-\varepsilon},\ \text{ if } m \ge 3 \text{ and } P^m H^p < \infty \text{ for all } p.$$

In the last expression, $\varepsilon > 0$ *can be arbitrarily small*[10]*.*
(b) If, additionally, the integrability conditions imposed on function H *also*

[10] Results for other combinations of m and p can be easily deduced from the proof. We omit them for brevity.

hold for functions H_{ω_m}, for all ω_m; and for all θ_1, θ_2, and for all s, $1 \le s \le \frac{m}{2}$, in the δ_0-neighborhood of 0,

$$P^{m-2s}\left[\left(h_{\theta_1}^{[m-2s]} - h_{\theta_2}^{[m-2s]}\right)^2\right] \le C\left\|\theta_1 - \theta_2\right\|,$$

then inequality (32) also holds (with the same rates a_n) with $U_n^{(m)}$ changed to $\hat{U}_n^{(m)}$.

5.2 Proofs of Theorems in Section 2

5.2.1 Asymptotic Normality and Consistency of the Bootstrap

Only the proof of Theorem 2 is provided. The proof of Theorem 1 is analogous (and simpler), and is close to the proofs in Sherman (1993) and Arcones, Giné and Chen (1994).

First, we obtain a quadratic approximation to the bootstrap objective function $\hat{U}_n h_\theta$. Define $\tau_\theta = P^{m-1}h_\theta$ and $A = -P\left[\partial^2\tau_0\right]$. By Assumptions 1 and 3, A is a symmetric, positive definite matrix; $P\left[\partial\tau_0\right] = 0$ (this is the first-order condition in the population maximization problem), and $P\left\|\partial\tau_0\right\|^2 < \infty$. Define

$$R_\theta(z) = \left[P^m h_\theta + m\pi_{1,m}h_\theta\right](z) - m\theta'\partial\tau_0(z) + \frac{1}{2}\theta' A\theta.$$

Using this and the Hoeffding decomposition for the bootstrapped U-statistic, we obtain

$$\hat{U}_n^{(m)}h_\theta = \theta'\hat{W}_n - \frac{1}{2}\theta' A\theta + \hat{\zeta}_{n,\theta} \tag{33}$$

where $\hat{W}_n = m\hat{P}_n\partial\tau_0$, and $\hat{\zeta}_{n,\theta}$ is the remainder:

$$\hat{\zeta}_{n,\theta} = \hat{P}_n R_\theta + \sum_{k=2}^{m}\binom{m}{k}\hat{U}_n^{(k)}\left(\pi_{k,m}h_\theta\right). \tag{34}$$

Let $\delta_0 > 0$ be such that the neighborhood $\mathcal{N}$ in Assumption 3 contains the ball of radius δ_0 with the center at zero. By Assumptions 3 (i), (ii), conditions $h_0 \equiv 0$, $P\partial\tau_0 = 0$, and the second-order Taylor expansion around zero,

$$\left|\hat{P}_n R_\theta\right| \le m\left(PM + \hat{P}_n M\right)\left\|\theta\right\|^3 + m\left\|\left(\hat{P}_n - P\right)\partial^2\tau_0\right\|\left\|\theta\right\|^2 \tag{35}$$

for all $\|\theta\| \leq \delta_0$.

Now we check conditions of Theorem 6. By the bootstrap Hoeffding decomposition, Assumptions 2, 5 and Lemma 8 (c),

$$P \sup_{\theta \in \Theta} \left| \hat{U}_n^{(m)} h_\theta - P^m h_\theta \right| \to 0,$$

which together with the identification assumption 1 implies consistency of $\hat{\theta}_n$ for 0. Clearly, $\hat{W}_n = O_p\left(n^{-1/2}\right)$. Next, use (35) and integrability conditions imposed in Assumption 3 to argue that $\hat{P}_n R_\theta$ satisfies condition (25), actually, the stronger condition

$$\sup_{\|\theta\| \leq \delta_n} \frac{\left| \hat{P}_n R_\theta \right|}{\|\theta\|^2} \to^p 0$$

whenever $\delta_n \to +0$.It is enough to show that $\hat{P}_n M = O_p(1)$ and $\left(\hat{P}_n - P\right) \partial^2 \tau_0 = o_p(1)$ under conditions $PM < \infty$ and $P \|\partial^2 \tau_0\| < \infty$. Both follow from the following: if $P|f| < \infty$ then $\left| \hat{P}_n f - Pf \right| = o_p(1)$. In fact, $\left(\hat{P}_n - P\right) f = \left(\hat{P}_n - P_n\right) f + (P_n - P) f$. The second term is $o_p(1)$ by the Law of Large Numbers, and the first term is $o_p(1)$ by the bootstrap weak law of large numbers given e.g. in Theorem 3.5 in Giné and Zinn (1990). Condition $P|f| < \infty$ is sufficient for condition (i) of that theorem. Then, $\hat{P} \left| \left(\hat{P}_n - P_n\right) f \right| = o_p(1)$. By the Chebyshev inequality, for any $\varepsilon > 0$, $\hat{P} \left\{ \left| \left(\hat{P}_n - P_n\right) f \right| > \varepsilon \right\} = o_p(1)$. The left-hand side is bounded by 1. Integrate over P to obtain

$$P \left\{ \left| \left(\hat{P}_n - P_n\right) f \right| > \varepsilon \right\} = o(1).$$

It remains to verify condition (25) for higher-order U-processes in (34). Use the maximal inequality, Lemma 8 (c) (with $p = 1$). For $k \geq 3$,

$$P \sup_{\theta \in \Theta} \left| \hat{U}_n^{(k)} (\pi_{k,m} h_\theta) \right| = O\left(n^{-3/2}\right).$$

For $k = 2$, take a sequence $\delta_n \to +0$ and apply the maximal inequality from Lemma 8 (d) to classes $\mathcal{H}_n = \{\pi_{2,m} h_\theta : \|\theta\| \leq \delta_n\}$:

$$P \sup_{\|\theta\| \leq \delta_n} \left| \hat{U}_n^{(2)} (\pi_{2,m} h) \right| = o\left(n^{-1}\right).$$

Conclude that condition (25) is satisfied for all $k \geq 2$.

By Theorem 6,

$$n^{1/2}\left(\hat{\theta}_n - \hat{P}_n A^{-1}\partial\tau_0\right) = o_p(1).$$

A similar derivation for the sample problem gives

$$n^{1/2}\left(\theta_n - P_n A^{-1}\partial\tau_0\right) = o_p(1).$$

Therefore,

$$\nu_n \equiv n^{1/2}\left(\hat{\theta}_n - \theta_n\right) - n^{1/2}\left(\hat{P}_n - P_n\right) A^{-1}\partial\tau_0 = o_p(1). \tag{36}$$

By Theorem 2.2 of Bickel and Freedman (1981), for almost all sequences $\{Z_1, Z_2, ...\}$

$$n^{1/2}\left(\hat{P}_n - P_n\right) A^{-1}\partial\tau_0 \to N(0, \Gamma).$$

Weak convergence to the multivariate normal distribution is always uniform (Corollary 2.6, Theorem 3.1 and Corollary 3.2 of Bhattacharya and Rao (1976)); therefore, for almost all sequences $\{Z_1, Z_2, ...\}$,

$$\sup_{A\in\mathcal{A}}\left|\int_A d\hat{F}_{n^{1/2}\left(\hat{P}_n - P_n\right)A^{-1}\partial\tau_0} - \int_A d\Phi_\Gamma\right| \to 0. \tag{37}$$

This and (36) imply the conclusion of Theorem 2:

$$\sup_{A\in\mathcal{A}}\left|\int_A d\hat{F}_{n^{1/2}\left(\hat{\theta}_n - \theta_n\right)} - \int_A d\Phi_\Gamma\right| = o_p(1), \tag{38}$$

as follows. For $\varepsilon > 0$, and a set $A \in \mathcal{A}$, define $A^\varepsilon = \cup\{B(x,\varepsilon), x \in A\}$, where $B(x,\varepsilon)$ is the open ball with center x and radius ε, and $A^{-\varepsilon} = \mathbb{R}^d \backslash \left(\mathbb{R}^d \backslash A\right)^\varepsilon$. Both sets are in $\mathcal{A}$ (both are convex, the first is open and the second is closed, so both are measurable), and $A^{-\varepsilon} \subset A \subset A^\varepsilon$. It is known that

$$\sup_{A\in\mathcal{A}}\int_{A^\varepsilon\backslash A^{-\varepsilon}} d\Phi_\Gamma \leq K(d,\Gamma)\,\varepsilon,$$

see formula (3) and Corollary 3.2 in Bhattacharya and Rao (1976). We have

$$\int_A d\hat{F}_{n^{1/2}\left(\hat{\theta}_n - \theta_n\right)} \leq \int_{A^\varepsilon} d\hat{F}_{n^{1/2}\left(\hat{W}_n - W_n\right)} + P\{\|\nu_n\| \geq \varepsilon\}$$

and

$$\int_A d\hat{F}_{n^{1/2}\left(\hat{\theta}_n-\theta_n\right)} \geq \int_{A^{-\varepsilon}} d\hat{F}_{n^{1/2}\left(\hat{W}_n-W_n\right)} - P\left\{\|\nu_n\| \geq \varepsilon\right\}.$$

Then,

$$\begin{aligned}
& \sup_{A\in\mathcal{A}} \left| \int_A d\hat{F}_{n^{1/2}\left(\hat{\theta}_n-\theta_n\right)} - \int_A d\Phi_\Gamma \right| \\
\leq\ & \sup_{A\in\mathcal{A}} \left| \int_A d\hat{F}_{n^{1/2}\left(\hat{P}_n-P_n\right)A^{-1}\partial\tau_0} - \int_A d\Phi_\Gamma \right| \\
& + \sup_{A\in\mathcal{A}} \int_{A^{\varepsilon}\setminus A^{-\varepsilon}} d\Phi_\Gamma + P\left\{\|\nu_n\| \geq \varepsilon\right\}.
\end{aligned}$$

Therefore, (38) holds.

5.2.2 Estimation of the Variance

Here we prove consistency of the asymptotic variance estimators given in Theorem 3. We consider only the bootstrap problem, while the (simpler) proof for the sample problem can be reconstructed using the same steps. We check conditions of part (b) of Theorem 6. By condition $P^m H^p < \infty$, for $p > 2$, the bootstrap Hoeffding decomposition, and Lemma 8 (c), for each $\varepsilon > 0$ there is $\eta > 0$ such that

$$\begin{aligned}
P\left\{\left\|\hat{\theta}_n\right\| > \varepsilon\right\} &\leq \eta^{-p} P \sup_{\theta\in\Theta} \left|\hat{U}_n^{(m)} h_\theta - P^m h_\theta\right|^p \\
&= \eta^{-p} O\left(n^{-p/2}\right) = o\left(n^{-1}\right).
\end{aligned}$$

Next note that conditions $P^m H^p < \infty$, $PM^p < \infty$, $P\left\|\partial^2\tau_0\right\|^p < \infty$, and the Taylor expansion, imply that $P\left\|\partial\tau_0\right\|^p < \infty$. Then by Rosenthal inequality, $P\left\|\hat{W}_n\right\|^p = O\left(n^{-p/2}\right)$, and, therefore,

$$\begin{aligned}
& nP\left\|\hat{W}_n\right\|^2 1\left\{\left\|\hat{W}_n\right\|^2 > \varepsilon\right\} \\
\leq\ & \frac{1}{\left(n\varepsilon^2\right)^{p-2}} P\left\|\hat{W}_n\right\|^p = o\left(n^{-1}\right).
\end{aligned}$$

The extra integrability assumptions of Theorem 3 ensure that $\hat{P}_n R_{n,\theta}$ satisfies condition (26). To check (26) for the higher-order U-processes, invoke Lemma

8 (c) with $p > 2$. Theorem 3 implies

$$P \left\| n^{1/2} \left(\hat{\theta}_n - A^{-1} \hat{P}_n \partial \tau_0 \right) \right\|^2 \to 0.$$

By Chebyshev inequality,

$$\hat{P} \left\| n^{1/2} \left(\hat{\theta}_n - A^{-1} \hat{P}_n \partial \tau_0 \right) \right\|^2 \to^p 0.$$

By Theorem 2.2 of Bickel and Freedman (1981),

$$\widehat{Var} \left(n^{1/2} A^{-1} \hat{P}_n \partial \tau_0 \right) - \Gamma \to^{a.s.} 0$$

so,

$$\widehat{Var} \left(n^{1/2} \hat{\theta}_n \right) - \Gamma \to^p 0.$$

5.2.3 Generic Bound for Rank Estimators

Here we prove Theorem 4 for the bootstrap problem. The proof for the sample problem follows the same steps. We use the same representation (33), but check conditions of Theorem 7. The rate in Theorem 4, a_n, is determined by the rate of convergence to zero of the U-process of order 2 in the remainder $\hat{\zeta}_{n,\theta}$, $\hat{U}_n^{(2)} (\pi_{2,m} h_\theta)$. It is given in Lemma 9. To apply it, consider the class of functions $\left\{ \breve{h} = \pi_{2,m} h \right\}$. The class consists of P-degenerate functions of two arguments. Note that by Jensen inequality, the condition on $P^{m-2} h_\theta$, $P^{m-2} h_\theta^{[m-2]}$ in Assumptions 6, 7 imply the same condition for functions $\breve{h}_\theta$, $\breve{h}_\theta^{[m-2]}$. If the class $\{h\}$ is Euclidean, then so is the class $\left\{ \breve{h} \right\}$ (see the properties of the Euclidean classes below). Also, the class $\left\{ \breve{h} \right\}$ inherits from the class $\{h\}$ its integrability properties (finiteness of moments). Lemma 9 (b) gives the rate, a_n, with which condition (iii) of Theorem 7 is satisfied for $\hat{U}_n^{(2)} (\pi_{2,m} h_\theta)$: $a_n = \left(n^{1/6} (\log n)^{-2/3} \right)^{1/(1+2/3p)}$ if $P^m H_{\omega_m}^p < \infty$ for $p \geq 6$ and all permutations ω_m. It now suffices to check that the other conditions of Theorem 7 are satisfied with this rate and the probability $1 - O\left(n^{-1/6}\right)$.

First, check condition (i). For $\hat{\theta}_n$, as in the previous subsection, for $p = 6$ (this is the minimal integrability assumption imposed in Theorem 4)

$$\begin{aligned} P \left\{ \left\| \hat{\theta}_n \right\| > \delta \right\} &\leq \eta_\delta^{-p} P \sup_{\theta \in \Theta} \left| \hat{U}_n^{(m)} h_\theta - P^m h_\theta \right|^p \\ &= \eta_\delta^{-p} O\left(n^{-p/2}\right) = O\left(n^{-3}\right). \end{aligned}$$

Since by the Rosenthal inequality, $P\left\|\hat{W}_n\right\|^4 = O\left(n^{-2}\right)$ under condition $P\left\|\partial\tau_0\right\|^4 < \infty$,

$$P\left\{\left\|\hat{W}_n\right\| > \delta\right\} = O\left(n^{-2}\right).$$

Condition (ii) is trivial here because A is assumed to be a constant positive definite matrix. As a consequence, $\hat{\eta}_n = \hat{W}_n$ except on an event of probability $O\left(n^{-2}\right)$. We, therefore, can neglect the distinction between $\hat{\eta}_n$ and $\hat{W}_n$.

Condition (iii) for higher-order U-processes in $\hat{\zeta}_{n,\theta}$, $\hat{U}_n^{(k)}\left(\pi_{k,m}h_\theta\right)$ (for $a_n \leq n^{1/6}$) is trivial because $\sup_\Theta\left|\hat{U}_n^{(k)}\left(\pi_{k,m}h_\theta\right)\right| = O_p\left(n^{-k/2}\right)$ by Lemma 8 (c), and the rate $n^{k/2}$, $k \geq 3$, dominates the rate na_n^2, which is at most $n^{4/3}$.

Condition (iii) for $\hat{P}_nR_\theta$ can be checked using[the same methods as in the previous proofs and] the extra integrability assumptions on $M(z)$, $\partial^2\tau_0$, and $\partial\tau_0$ made in Theorem 4. We, therefore, have

$$P\left\{\left\|n^{1/2}\left(\hat{\theta}_n - A^{-1}\hat{P}_n\partial\tau_0\right)\right\| > Ka_n^{-1}\right\} = O\left(a_n^{-1}\right).$$

A similar derivation gives

$$P\left\{\left\|n^{1/2}\left(\theta_n - A^{-1}P_n\partial\tau_0\right)\right\| > Ka_n^{-1}\right\} = O\left(a_n^{-1}\right).$$

Therefore,

$$\begin{aligned} & P\left\{\nu_n \equiv \left\|n^{1/2}\left(\hat{\theta}_n - \theta_n\right) - n^{1/2}A^{-1}\left(\hat{P}_n - P_n\right)\partial\tau_0\right\| > Ka_n^{-1}\right\} \quad (39) \\ = & \; O\left(a_n^{-1}\right) \end{aligned}$$

for some $K > 0$.

Next we use the multivariate Berry-Esséen Theorem (Corollary 18.3 in Bhattacharya and Rao (1976)). For the sample problem, under conditions that $Var\left(\partial\tau_0\right)$ is a positive definite matrix and $P\left\|\partial\tau_0\right\|^3 < \infty$, we have:

$$\sup_{A\in\mathcal{A}}\left|\int_A dF_{A^{-1}P_n\partial\tau_0} - \int_A d\Phi_\Gamma\right| \leq n^{-1/2}c\left(d\right)P\left\|\Gamma^{-1/2}A^{-1}\partial\tau_0\right\|^3,$$

where $c\left(d\right)$ is an absolute constant for each d.

For the bootstrap problem, let C_0 be a constant such that

$$\limsup_{n\to\infty} P_n\left\|\Gamma_n^{-1/2}A^{-1}\partial\tau_0\right\|^3 < C_0(P-a.s),$$

where $\Gamma_n = \widehat{Var}\left(n^{1/2} A^{-1} \hat{P}_n \partial\tau_0\right)$. Such finite constant exists by the law of large numbers under conditions that Γ, A are positive definite and $P \left\|\partial\tau_0\right\|^3 < \infty$. Apply the Berry-Esséen Theorem conditionally on sequences of data for which this condition is satisfied. Then, $P - a.s.$,

$$\lim_{n\to\infty} \sup_{A\in\mathcal{A}} \left| \int_A d\hat{F}_{A^{-1}\left(\hat{P}_n - P_n\right)\partial\tau_0} - \int_A d\Phi_{\Gamma_n} \right| \leq n^{-1/2} c(d)\, C_0$$

Integrate over P and take into account that the integrand is a sequence of bounded functions, apply the Lebesgue dominated convergence theorem:

$$\lim_{n\to\infty} P \sup_{A\in\mathcal{A}} \left| \int_A d\hat{F}_{A^{-1}\left(\hat{P}_n - P_n\right)\partial\tau_0} - \int_A d\Phi_{\Gamma_n} \right| \leq n^{-1/2} c(d)\, C_0$$

or, by the Chebyshev inequality,

$$\sup_{A\in\mathcal{A}} \left| \int_A d\hat{F}_{A^{-1}\left(\hat{P}_n - P_n\right)\partial\tau_0} - \int_A d\Phi_{\Gamma_n} \right| = O_p\left(n^{-1/2}\right).$$

Finally, condition $P \left\|\partial\tau_0\right\|^4 < \infty$, implies that $\Gamma_n - \Gamma = O_p\left(n^{-1/2}\right)$. Then it follows from the properties of the normal distribution that

$$\sup_{A\in\mathcal{A}} \left| \int_A d\Phi_{\Gamma_n} - \int_A d\Phi_{\Gamma} \right| = O_p\left(n^{-1/2}\right).$$

So, we have

$$\sup_{A\in\mathcal{A}} \left| \int_A d\hat{F}_{A^{-1}\left(\hat{P}_n - P_n\right)\partial\tau_0} - \int_A d\Phi_{\Gamma} \right| = O_p\left(n^{-1/2}\right) \tag{40}$$

Now we obtain the uniform result of Theorem 4. We show it for the bootstrap. Use (39) and (40), and the logic of the proof of uniformity in consistency theorems. Let $\varepsilon_n = K a_n^{-1}$. We have:

$$\begin{aligned}
& \sup_{A\in\mathcal{A}} \left| \int_A d\hat{F}_{n^{1/2}\left(\hat{\theta}_n - \theta_n\right)} - \int_A d\Phi_{\Gamma} \right| \\
\leq\; & \sup_{A\in\mathcal{A}} \left| \int_A d\hat{F}_{n^{1/2}\left(\hat{P}_n - P_n\right)A^{-1}\partial\tau_0} - \int_A d\Phi_{\Gamma} \right| \\
& + \sup_{A\in\mathcal{A}} \int_{A^{\varepsilon_n}\setminus A^{-\varepsilon_n}} d\Phi_{\Gamma} + P\left\{\left\|\nu_n\right\| \geq \varepsilon_n\right\} \\
=\; & O_p\left(n^{-1/2}\right) + O\left(\varepsilon_n\right) + O_p\left(a_n^{-1}\right) = O_p\left(a_n^{-1}\right).
\end{aligned}$$

5.2.4 Better Rates Under Additional Smoothness Assumptions

Under additional Assumption 8, the degenerate U-processes of order up to $s \geq 2$ in the Hoeffding decomposition of the criterion function $G_{n,\theta}$ are all smooth functions of θ. Then one can approximate θ_n by the random vector η_n which solves the problem

$$\eta_n \in \arg\max_{\theta\in\Theta} G^0_{n,\theta} \equiv U_n^{(s)} h^*_\theta$$

where

$$h^*_\theta = \sum_{k=0}^{s} \binom{m}{k} \pi_{k,m} h_\theta = \sum_{k=0}^{s} \binom{m}{k} \pi_{k,s} f_\theta$$

The bootstrapped estimator, $\hat{\theta}_n$ can be approximated by

$$\hat{\eta}_n \in \arg\max_{\theta\in\Theta} \hat{U}_n^{(s)} h^*_\theta.$$

The properties of η_n and $\hat{\eta}_n$ can be found by powerful methods based on the Taylor expansion and Berry-Esséen bounds for higher-order U-statistics. Note first that by the Hoeffding decomposition, maximal and Chebyshev inequalities, for any $\delta > 0$,

$$P\left\{\|\eta_n\| > \delta\right\} = O\left(n^{-1/2}\right),$$

and

$$P\left\{\|\hat{\eta}_n\| > \delta\right\} = O\left(n^{-1/2}\right).$$

In particular, with probability at least $1-$ $O\left(n^{-1/2}\right)$, $\eta_n - \theta_0$ coincides with the solution to the first order condition:

$$U_n^{(2)} g_{\theta_0+\theta} = \xi_{n,\theta_0+\theta}$$

where $g_\theta = \left(P^m + m\pi_{1,s} + \frac{m(m-1)}{2}\pi_{2,s}\right)\partial f_\theta$, and $\xi_{n,\theta} = \sum_{k=3}^{s}\binom{m}{k} U_n^{(k)} \pi_{k,s} \partial f_\theta$. Functions g_θ, $\xi_{n,\theta}$, and η_n satisfy the assumptions of Lemma 17 (in particular, $P^m \partial g_0 = P^m \partial f_0 = 0$, by the first-order condition in the population problem), and, therefore, the following Berry-Esséen bound holds:

$$\sup_{A\in\mathcal{A}} \left| \int_A dF_{n^{1/2}(\eta_n - \theta_0)} - \int_A d\Phi_\Gamma \right| = O\left(n^{-1/2}\right) \qquad (41)$$

To obtain a similar bound for the bootstrap problem, apply Lemma 17 conditionally on the sample. With probability at least $1- O\left(n^{-1/2}\right)$, the random vector $\hat{\eta}_n - \eta_n$ coincides with the solution to the equation:

$$\hat{U}_n^{(2)} g_{\eta_n+\theta} = \hat{\xi}_{n,\eta_n+\theta}$$

where $\hat{\xi}_{n,\theta} = \sum_{k=3}^{s} \binom{m}{k} \hat{U}_n^{(k)} \pi_{k,s} \partial f_\theta$. Note, in particular, that

$$\begin{aligned} P_n^2 g_{\eta_n+\theta} &= \frac{n-1}{n} U_n g_{\eta_n} + \frac{1}{n^2} \sum_{i=1}^{n} g_{\eta_n}\left(Z_i, Z_i\right) \\ &= \xi_{n,\eta_n} + \frac{1}{n^2} \sum_{i=1}^{n} g_{\eta_n}\left(Z_i, Z_i\right); \end{aligned}$$

and, therefore, satisfies Assumption (ii) of Lemma 17 with $K = O_p(1)$. Also note that the conditional moments required to apply Lemma 17 are bounded for almost all sequences of data $\{Z_1, Z_2 ...\}$, by the moment conditions on L in Assumption 10 (a); therefore, c_d in Lemma 17 will be $O_p(1)$. Thus,

$$\sup_{A \in \mathcal{A}} \left| \int_A d\hat{F}_{n^{1/2}(\hat{\eta}_n - \eta_n)} - \int_A d\Phi_{\Gamma_n} \right| = O_p\left(n^{-1/2}\right).$$

Under the assumption that $P \left\|\partial f_\theta\right\|^4 < \infty$, we can rewrite the last bound as

$$\sup_{A \in \mathcal{A}} \left| \int_A d\hat{F}_{n^{1/2}(\hat{\eta}_n - \eta_n)} - \int_A d\Phi_{\Gamma} \right| = O_p\left(n^{-1/2}\right). \tag{42}$$

The objective function for the estimators θ_n, $\hat{\theta}_n$, contains additional terms given by:

$$\zeta_{n,\theta} = \sum_{k=s+1}^{m} \binom{m}{k} U_n^{(k)}\left(\pi_{k,m} h_\theta\right) + r_{n,\theta}.$$

and

$$\hat{\zeta}_{n,\theta} = \sum_{k=s}^{m} \binom{m}{k} \hat{U}_n^{(k)}\left(\pi_{k,m} h_\theta\right) + \hat{r}_{n,\theta}$$

To estimate the differences $\theta_n - \eta_n$ and $\hat{\theta}_n - \hat{\eta}_n$, use Theorem 7. Under Assumption 8, by the Taylor expansion,

$$U_n^{(s)} h_\theta = U_n^{(s)} h_\theta^* = \theta' W_n - \frac{1}{2} \theta' A_{n,\theta} \theta$$

where

$$W_n = U_n^{(s)} \partial h_0^*$$

and, for $A = -\partial^2 P h_0 = -\partial^2 P h_0^*$,

$$\partial^2 G_{n,\theta}^0 = -A + U_n^{(s)} \partial^2 \left(h_0^* - P h_0^*\right) + O\left(L \|\theta\|\right)$$

Condition (i) of Theorem 7 is satisfied with $a_n = n^{1/2}$, for both θ_n and η_n. Condition (ii) follows from the previous display, positive definiteness of A, and the moment conditions on function L. If Assumptions 8, 9 are satisfied with $s = 2$, then condition (iii) of the Theorem is satisfied with $a_n = n^{1/2-\varepsilon}$, where $\varepsilon > 0$ is arbitrarily small, by Lemma 9 (a) (for the degenerate U-process of order 3), and 8 (a) with sufficiently high p (for the degenerate U-processes of order 4 and higher). If $s = 3$, then condition (iii) is satisfied with $a_n = n^{3/4-\varepsilon}$, by the same Lemmas (Lemma 9 now should be used for the degenerate U-processes of order 4)[11]. Conditions (i-iii) can be verified for the bootstrap (i.e. relative to the unconditional distribution of the bootstrap draws) in the same way. Particularly, condition (iii) directly follows from Lemmas 8 (c) and 9 (b).

It follows that for some constant $K > 0$,

$$P\left\{n^{1/2} \|\theta_n - \eta_n\| > K a_n^{-1}\right\} = O\left(a_n^{-1}\right) \tag{43}$$

and

$$P\left\{n^{1/2} \left\|\hat{\theta}_n - \hat{\eta}_n\right\| > K a_n^{-1}\right\} = O\left(a_n^{-1}\right).$$

Combining the last to bounds we have

$$P\left\{n^{1/2} \left\|\left(\hat{\theta}_n - \theta_n\right) - \left(\hat{\eta}_n - \eta_n\right)\right\| > K a_n^{-1}\right\} = O\left(a_n^{-1}\right). \tag{44}$$

The sample version of Theorem 5 follows from (41) and (43), while its bootstrap counterpart follows from (42) and (44).

[11] The result for $s = 3$ can be used to obtain the Edgeworth expansion for the distribution functions of $n^{1/2}\theta_n$ and $n^{1/2}\left(\hat{\theta}_n - \theta_n\right)$ with the error term of order $O\left(n^{-3/4+\varepsilon}\right)$ ($O_p\left(n^{-3/4+\varepsilon}\right)$ for the bootstrap), which implies that the *symmetric* confidence intervals for θ_n constructed using the bootstrap are $O_p\left(n^{-3/4+\varepsilon}\right)$-accurate. The last bound may not be tight, even with ε omitted. This is similar to the case of the parametric estimators, whose symmetric confidence intervals are also more accurate than the one-sided ones, and have the error of coverage probability $O_p\left(n^{-1}\right)$ (see Hall (1992)). The derivation is tedious because it requires further terms in the Edgeworth expansion for η_n and $\hat{\eta}_n$, and is omitted.

5.3 Bounds on Oscillations of U-Processes

Here we provide a brief discussion of the empirical process theory for U-processes, and extensions to it, that eventually lead to Lemmas 8, 9. The bounds listed here are relevant for the U-processes indexed by a Euclidean class of functions. For the convenience of the reader we remind the definition. Call function H an envelope of a class of functions $\mathcal{H}$ if $|h| \leq H$ for each $h \in \mathcal{H}$.

Definition 10 *(Nolan and Pollard (1987)) Let $\mathcal{H}$ be a class of real-valued functions defined on the same set. Call $\mathcal{H}$ Euclidean for the envelope H if there exist positive constants (referred to as Euclidean numbers in the sequel) A and V such that for any measure μ, for which $0 < \mu H < \infty$,*

$$N_2(\varepsilon, d_\mu) \leq A\varepsilon^{-V}, \; 0 < \varepsilon \leq 1.$$

Here, for $h_1, h_2 \in \mathcal{H}$, $d_\mu(h_1, h_2) = \mu|h_1 - h_2|^2/\mu H^2$ and $N_2(\varepsilon, d_\mu)$ is the packing number of $\mathcal{H}$ with respect to the pseudometric d_μ, i.e. the largest number N such that there exist functions $h_1, ..., h_N$ with the property $d_\mu(h_i, h_j) > \rho$ for $i \neq j$.

A detailed review of the properties of Euclidean classes of functions can be found in Nolan and Pollard (1987) and Pakes and Pollard (1989). In particular, if $\mathcal{H}_1$ and $\mathcal{H}_2$ are two Euclidean classes for envelopes, respectively, H_1 and H_2, then class $\mathcal{H}_1 + \mathcal{H}_2 \equiv \{h_1 + h_2 : h_i \in \mathcal{H}_i\}$ is Euclidean for envelope $H_1 + H_2$ and class $\mathcal{H}_1 \cdot \mathcal{H}_2 \equiv \{h_1 \cdot h_2 : h_i \in \mathcal{H}_i\}$ is Euclidean for envelope $H_1 \cdot H_2$. If $\mathcal{H} = \{h : \mathcal{Z}^m \to \mathbb{R}\}$ is A, V- Euclidean for envelope H, then the class $\{|h| : h \in \mathcal{H}\}$ is A, V-Euclidean for the envelope H, and for any probability distribution μ, acting on variables $z_1, ..., z_k$, class

$$\{\mu h(\cdot, z_{k+1}, ..., z_m) : h \in \mathcal{H}\}$$

is A, V-Euclidean for the envelope μH (in particular, μ may put mass 1 on a value of $(z_1, ..., z_k)$).

It is convenient to introduce extra notation for the rest of this subsection. Throughout $\lesssim$ will denote inequality up to a multiplicative constant. The constant may depend on certain parameters of the model (typically, the Euclidean numbers A and V, the order of the process m and so on), but not on n or the sample data $\{Z_1, ..., Z_n\}$. In particular, we will often use the inequality $(a + b)^p \lesssim a^p + b^p$, $a, b \geq 0$, $p > 0$, where the constant depends on p only (for $p \in (0,1)$ the constant is 1). Symbol $\|\cdot\|_{\mathcal{H}}$ will stand for the supremum over a class of functions $\mathcal{H}$.

Lemma 11 gives bounds for the first moment of the suprema of the degenerate empirical and U-processes.

Lemma 11 *Let $\mathcal{H}$ be a class of P-degenerate symmetric functions which is Euclidean for an envelope H with $P^m H > 0$. Then*

$$\begin{aligned} & P^{\infty} \left\| U_n^m h \right\|_{\mathcal{H}} \\ \lesssim \ & n^{-m/2} P^{\infty} \left[\left(U_n^m H^2 \right)^{1/2} \int_0^{\left(\left\| U_n^m h^2 \right\|_{\mathcal{H}} / U_n^m H^2 \right)^{1/2}} (1 - \log \varepsilon)^{m/2} \, d\varepsilon \right] \end{aligned}$$

Here the multiplicative constant depends on m and the Euclidean numbers A, V only.

Proof. Cases $m = 1, 2$ were considered in Pollard (1989), Theorem 4.2 (i), and Nolan and Pollard (1987). For $m \geqslant 1$, the inequalities follow from Propositions 2.1, 2.2 and 2.6 in Arcones and Giné (1993) (see also the calculations in (1994)). ■

Remark 12 *The integral that appears in Lemma 11 (with $\frac{\left\| U_n^m h^2 \right\|_{\mathcal{H}}}{U_n^m H^2} \equiv x \in (0,1]$) can be bounded from above and from below by multiples of function*

$$J_m(x) = x^{1/2} \left(1 - \frac{1}{m} \log x \right)^{m/2},$$

which is increasing, concave, and bounded on $x \in (0,1]$. Furthermore, $J_m(x)$, $m \geq 1$, satisfies

$$(m/2)^{m/2} (\log n)^{-m/2} J_m(x) \leq x^{1/2} \vee \left(n^{-1} \log n \right)^{1/2}$$

for all $x \in (0,1]$ and $n \geq e^m$ (particularly, if $x \leq n^{-1} \log n$, $J_m(x) \leq J(n^{-1} \log n)$ by monotonicity).

The bound on $P^{\infty} \left\| U_n^m h \right\|_{\mathcal{H}}$ is related to the "continuity modulus" of the class $\mathcal{H}$, $\left\| P^m h^2 \right\|_{\mathcal{H}}^{1/2}$.

Lemma 13 *Let $\mathcal{H} = \{h : \mathcal{Z}^m \to \mathbb{R}\}, m \geq 1,$ be a Euclidean class of symmetric, P-degenerate functions with envelope 1. Then for all n,*

$$P^\infty \left\|U_n^{(m)} h\right\|_{\mathcal{H}} \lesssim \left(n^{-1} \log n\right)^{m/2} \left\|P^m h^2\right\|_{\mathcal{H}}^{1/2} + \left(n^{-1} \log n\right)^{(m+1)/2}, \tag{45}$$

where the multiplicative constant depends on m and the Euclidean numbers of the class only.

Proof. Follows from Theorem 8 in Giné and Mason (2007). ■

Lemma 14 *Let $\mathcal{H} = \{h : \mathcal{Z}^m \to \mathbb{R}\}$ be a class of symmetric, P-degenerate functions, Euclidean for an envelope H. If for $p \geq 2$, $P^m H^p < \infty$, then*

$$P^\infty \left\|U_n^{(m)} h\right\|_{\mathcal{H}} \lesssim \left(n^{-1} \log n\right)^{m/2} \left\|P^m h^2\right\|_{\mathcal{H}}^{1/2} + \left(n^{-1} \log n\right)^{(m+1)/2-1/p}.$$

In these inequalities, the multiplicative constants depend on m, $P^m H^p$ and the Euclidean numbers of the class only.

Proof. First, we obtain

$$P^\infty \left\|U_n^{(m)} h^2\right\|_{\mathcal{H}} \lesssim \left\|P^m h^2\right\|_{\mathcal{H}} + \left(n^{-1} \log n\right)^{1-2/p}. \tag{46}$$

Let $\mathcal{H}_L$, $L \geq 1$, be the class of functions $\{h1\{|h| \leq L\} : h \in \mathcal{H}\}$. Note that $\mathcal{H}_L$ is Euclidean for the envelope L. Consider the case $L = 1$. By the Hoeffding decomposition, Lemma 11 and Remark 12,

$$\begin{aligned} P^\infty \left\|U_n^{(m)} h^2\right\|_{\mathcal{H}_1} &\lesssim \left\|P^m h^2\right\|_{\mathcal{H}_1} + n^{-1/2} P^\infty J_1\left(\left\|P_n\left(\pi_{1,m} h^2\right)^2\right\|_{\mathcal{H}_1}\right) + n^{-1} \\ &\lesssim \left\|P^m h^2\right\|_{\mathcal{H}_1} \\ &\quad + n^{-1/2} P^\infty \left\|P_n\left(\pi_{1,m} h^2\right)^2 \log n\right\|_{\mathcal{H}_1}^{1/2} + n^{-1} \log n \end{aligned}$$

Note that

$$\begin{aligned} \left\|P_n\left(\pi_{1,m} h^2\right)^2\right\|_{\mathcal{H}_1} &\lesssim \left\|P_n\left(P^{m-1} h^2\right)^2\right\|_{\mathcal{H}_1} + \left\|P^m h^4\right\|_{\mathcal{H}_1} \\ &\leq \left\|P_n P^{m-1} h^2\right\|_{\mathcal{H}_1} + \left\|P^m h^2\right\|_{\mathcal{H}_1} \end{aligned}$$

Therefore, (also using $2|xy| \leq x^2 + y^2$)

$$\begin{aligned} P^\infty \left\|U_n^{(m)} h^2\right\|_{\mathcal{H}_1} &\lesssim \left\|P^m h^2\right\|_{\mathcal{H}_1} + n^{-1} \log n \\ &\quad + \left(n^{-1} \log n\right)^{1/2} P^\infty \left\|P_n P^{m-1} h^2\right\|_{\mathcal{H}_1}^{1/2} \end{aligned} \tag{47}$$

Apply this inequality to the process $P_n(P^{m-1}h^2)$; denoting by X the expression $P^\infty \|P_n(P^{m-1}h^2)\|_{\mathcal{H}_1}$, and by $C > 0$ the multiplicative constant,

$$X \le C \left\|P^m h^2\right\|_{\mathcal{H}_1} + C X^{1/2} \left(n^{-1} \log n\right)^{1/2} + C n^{-1} \log n$$

One possibility is that $X > 4C^2 n^{-1} \log n$, in which case the previous inequality gives

$$X \le C \left\|P^m h^2\right\|_{\mathcal{H}_1} + \frac{1}{2} X + C n^{-1} \log n,$$

so that

$$X \lesssim \left\|P^m h^2\right\|_{\mathcal{H}_1} + n^{-1} \log n.$$

The other possibility is that $X \ge 4C^2 n^{-1} \log n$. In both cases,

$$P^\infty \left\|P_n\left(P^{m-1}h^2\right)\right\|_{\mathcal{H}_1} \le \left\|P^m h^2\right\|_{\mathcal{H}_1} + n^{-1} \log n.$$

Substitute this into (47):

$$P^\infty \left\|U_n^{(m)} h^2\right\|_{\mathcal{H}_1} \lesssim \left\|P^m h^2\right\|_{\mathcal{H}_1} + n^{-1} \log n.$$

For an arbitrary $L \ge 1$, by rescaling,

$$P^\infty \left\|U_n^{(m)} h^2\right\|_{\mathcal{H}_L} \lesssim \left\|P^m h^2\right\|_{\mathcal{H}_L} + L^2 n^{-1} \log n.$$

Next, as

$$\begin{aligned} h^2 &= h^2 1\{|h| \le L\} + h^2 1\{|h| \ge L\} \\ &\le h^2 1\{|h| \le L\} + H^2 1\{H \ge L\}, \end{aligned}$$

we have

$$\begin{aligned} & P^\infty \left\|U_n^{(m)} h^2\right\|_{\mathcal{H}} \\ \lesssim\ & \left\|P^m h^2\right\|_{\mathcal{H}} + L^2\left(n^{-1} \log n\right) + P H^2 1\{H > L\} \\ =\ & \left\|P^m h^2\right\|_{\mathcal{H}} + L^2\left(n^{-1} \log n\right) + o\left(L^{-p+2}\right). \end{aligned}$$

Taking $L = \left(n^{-1} \log n\right)^{-1/p}$ gives (46).

For a U-statistic of order m, use first Lemma 11 and Remark 12:

$$\begin{aligned} P^\infty \left\|U_n^{(m)} h\right\|_{\mathcal{H}} \lesssim\ & \left(n^{-1} \log n\right)^{m/2} P^\infty \left\|U_n^{(m)} h^2\right\|_{\mathcal{H}} \\ & + \left(P^m H^2\right)^{1/2} \left(n^{-1} \log n\right)^{(m+1)/2}. \end{aligned}$$

Now use (46). ■

Lemma 15 *(Hoffmann-Jørgensen inequality for U-Processes indexed by Euclidean classes of functions). Let $\mathcal{H} = \{h : \mathcal{Z}^m \to \mathbb{R}\}$ be a class of P-degenerate symmetric functions which is Euclidean for a P^m-square-integrable envelope H. Then for every $p \geq 2$*

$$P^\infty \left\|U_n^{(m)} h\right\|_{\mathcal{H}}^p \lesssim \left(P^\infty \left\|U_n^{(m)} h\right\|_{\mathcal{H}}\right)^p + n^{-p(m+1)/2+1} P^m H^p,$$

with a constant depending on m, p and Euclidean constants A, V of the class only.

Proof. For $m = 1$ this inequality is well-known: it holds without constraints on the capacity of the class $\mathcal{H}$, see van der Vaart and Wellner (1996), Theorem 2.14.5. For $m \geq 2$, Giné and Zinn (1992), Corollary 4, obtained the following bound (also without capacity restrictions on $\mathcal{H}$):

$$\begin{aligned} P^\infty \left\|U_n^{(m)} h\right\|_{\mathcal{H}}^p &\lesssim \left(P^\infty \left\|U_n^{(m)} h\right\|_{\mathcal{H}}\right)^p \\ &+ P^\infty \max_{i_m \leq n} \left\| \binom{n}{m}^{-1} \sum_{i_1,\dots,i_{m-1}:(i_1,\dots,i_{m-1}) \in I_n^{(m)}}^{n} h(Z_{i_1}, \dots, Z_{i_m}) \right\|_{\mathcal{H}}^p. \end{aligned}$$

The second term can be bounded by

$$\begin{aligned} & P^\infty \sum_{i_m=1}^{n} \left\| \binom{n}{m}^{-1} \sum_{i_1,\dots,i_{m-1}:(i_1,\dots,i_{m-1}) \in I_n^{(m)}}^{n} h(Z_{i_1}, \dots, Z_{i_m}) \right\|_{\mathcal{H}}^p \\ \lesssim\ & n^{-p+1} P' P^\infty \left\| U_{n-1}^{(m-1)} h(\cdot, Z') \right\|_{\mathcal{H}}^p, \end{aligned}$$

where Z' is an independent copy of Z_i, and P' integrates over Z'. Using the same argument for $U_{n-1}^{(m-1)} h(\cdot, z)$, we have, for each fixed z:

$$\begin{aligned} P^\infty \left\| U_{n-1}^{(m-1)} h(\cdot, z) \right\|_{\mathcal{H}}^p \ \lesssim\ & \left(P^\infty \left\| U_{n-1}^{(m-1)} h(\cdot, z) \right\|_{\mathcal{H}} \right)^p \\ & + n^{-p+1} P' P^\infty \left\| U_{n-2}^{(m-2)} h(\cdot, Z', z) \right\|_{\mathcal{H}}^p. \end{aligned}$$

Euclidean property of the class $\mathcal{H}$ gives an upper bound for the first term:

$$\left(P^\infty \left\| U_{n-1}^{(m-1)} h(\cdot, z) \right\|_{\mathcal{H}} \right)^p \lesssim n^{-(m-1)p/2} \left(PH(\cdot, z)^2 \right)^{p/2}$$

where the multiplicative constant is the same for all z.

Continue by induction, and use eventually the Hoffmann-Jørgensen inequality for $m = 1$ for the remaining P-process:

$$\begin{aligned} P^{\infty} \left\| U_n^{(m)} h \right\|_{\mathcal{H}}^p &\lesssim \left(P^{\infty} \left\| U_n^{(m)} h \right\|_{\mathcal{H}} \right)^p + \sum_{s=1}^{m-1} n^{(-p+1)s-(m-s)p/2} P^m H^p \\ &\quad + n^{(-p+1)(m-1)} n^{-1+1/p} P^m H^p \\ &\lesssim \left(P^{\infty} \left\| U_n^{(m)} h \right\|_{\mathcal{H}} \right)^p + n^{-(m+1)p/2+1} P^m H^p. \end{aligned}$$

■

Now consider the bootstrap version of the U-process. As in the preceding literature (e.g. Theorem 2.2 in Arcones and Giné (1994)) the goal is to relate the moments of the bootstrapped process $\hat{U}_n^{(m)} h$ to moments of a modified sample process, by using the symmetrization and poissonization techniques suggested in Giné and Zinn (1990). Note, however, that decomposition (30) requires the result under the assumption that h is P-degenerate, rather than P_n-degenerate, as it was assumed by previous authors.

We need extra notation. Let $Q_i^{(j)}, i = 1, 2, ...; j = 1, ..., m$, be i.i.d. (across i and j) random variables, independent of all Z_i, and having the Poisson distribution with parameter 1/2. Define random vectors

$$\tilde{Z}_i = \left(Z_i, Q_i^{(1)}, ..., Q_i^{(m)} \right);$$

let $\tilde{P}$ be the distribution of each $\tilde{Z}_i$, and $\tilde{h} \left(\tilde{z}_1, ..., \tilde{z}_m \right)$ be a symmetrized version of the function

$$\tilde{h}^0 \left(\tilde{z}_1, ..., \tilde{z}_m \right) = h \left(z_1, ..., z_m \right) q_1^{(1)} \cdot ... \cdot q_m^{(m)},$$

where $\tilde{z} = \left(z, q^{(1)}, ..., q^{(m)} \right)$. Note that functions $\tilde{h}$ are degenerate relative to the distribution $\tilde{P}$. The usefulness of the following lemma stems from the fact that the class of functions $\tilde{\mathcal{H}} = \left\{ \tilde{h} : h \in \mathcal{H} \right\}$ inherits the capacity and integrability properties (relative to $\tilde{P}$) from those of the class $\mathcal{H}$. In particular, if $\mathcal{H}$ is Euclidean for an envelope $H \left(z_1, ..., z_m \right)$, then $\tilde{\mathcal{H}}$ is Euclidean for a symmetrized version of the envelope $H \left(z_1, ..., z_m \right) q_1^{(1)} \cdot ... \cdot q_m^{(m)}$, denoted $\tilde{H}$. Also, since all moments of $Q_i^{(j)}$ are finite, $\{Z_i\}$ and $\left\{ Q_i^{(j)} \right\}$ are independent, $\tilde{H}$ has as many finite moments relative to $\tilde{P}$, as H does relative to P.

Lemma 16 *Let $\mathcal{H} = \{h : \mathcal{Z}^m \to \mathbb{R}\}$ be a class of P-degenerate real symmetric functions. Assume that $\mathcal{H}$ has an envelope H, and $P^m H^p < \infty$. Then*

$$P \left\| \hat{U}_n^{(m)} h \right\|_{\mathcal{H}}^p \lesssim P \left\| \frac{1}{n^m} \sum_{i_1,...,i_m} \tilde{h}\left(\tilde{Z}_{i_1}, ..., \tilde{Z}_{i_m}\right) \right\|_{\tilde{\mathcal{H}}}^p$$

where the constant depends on m and p only.

Proof. Use Hoeffding decomposition of the bootstrapped statistic relative to P_n (i.e. conditionally on the sample):

$$\hat{U}_n^{(m)} h = \sum_{k=0}^{m} \binom{m}{k} \hat{U}_n^{(k)} \left(\pi_{k,m}^{P_n} h\right)$$

where

$$\left(\pi_{k,m}^{P_n} h_\theta\right)(z_1, ..., z_k) = (\delta_{z_1} - P_n) ... (\delta_{z_k} - P_n) P_n^{m-k} h_\theta.$$

Next we show that for each $k = 0, ..., m$,

$$\begin{aligned} P \left\| \hat{U}_n^{(k)} \left(\pi_{k,m}^{P_n} h\right) \right\|_{\mathcal{H}}^p &\lesssim P \left\| \frac{1}{n^m} \sum_{i_1,...,i_m} \tilde{h}^0 \left(\tilde{Z}_{i_1}, ..., \tilde{Z}_{i_m}\right) \right\|_{\tilde{\mathcal{H}}}^p \\ &= P \left\| \frac{1}{n^m} \sum_{i_1,...,i_m} \tilde{h}\left(\tilde{Z}_{i_1}, ..., \tilde{Z}_{i_m}\right) \right\|_{\tilde{\mathcal{H}}}^p \end{aligned}$$

(the last equality is immediate).

Denote by E the expectation conditional on the sample $Z_1, ..., Z_n$. Let $\left\{\hat{Z}_1^{(j)}, ..., \hat{Z}_n^{(j)}\right\}$ be i.i.d. samples from P_n, independent across $j = 1, ..., k$; denote by $\hat{P}_n^{(j)}$ the bootstrap empirical measure that puts mass $1/n$ on each $\hat{Z}_i^{(j)}$. Let $N_1^{(j)}, ..., N_n^{(j)}$ be i.i.d. across i and j, independent from all Z_i, $\hat{Z}_i^{(j)}$, and each distributed as a difference between two independent Poisson r.v.

with parameter 1/2. Then

$$
\begin{aligned}
& E \left\| \hat{U}_n^{(k)} \left(\pi_{k,m}^{P_n} h \right) \right\|_{\mathcal{H}}^p \\
\lesssim \; & E \left\| \frac{1}{n^k} \sum_{i_1,\dots,i_k \ distinct} \left(\pi_{k,m}^{P_n} h \right) \left(\hat{Z}_{i_1}^{(1)}, \dots, \hat{Z}_{i_k}^{(k)} \right) \right\|_{\mathcal{H}}^p \\
\lesssim \; & E \left\| \frac{1}{n^k} \sum_{i_1,\dots,i_k} \left(\pi_{k,m}^{P_n} h \right) \left(\hat{Z}_{i_1}^{(1)}, \dots, \hat{Z}_{i_k}^{(k)} \right) \right\|_{\mathcal{H}}^p \\
= \; & E \left\| \left(\hat{P}_n^{(1)} - P_n \right) \left(\hat{P}_n^{(k)} - P_n \right) P_n^{m-k} h \right\|_{\mathcal{H}}^p =: (*)
\end{aligned}
$$

Here the first inequality follows by the decoupling inequality of de la Peña (1992), applied conditionally on $Z_1, \dots, Z_n$. In the second inequality the LHS is different from the RHS in that the latter includes summation over coinciding indices $i_1, \dots, i_k$. The second inequality follows from the following observation: for any r.v. X_h, Y_h, if $E[Y_h|X_h] = 0$, then, by the convexity inequality, $E\|X_h + Y_h\|_{\mathcal{H}}^p \geq E\|X_h + E[Y_h|X_h]\|_{\mathcal{H}}^p = E\|X_h\|_{\mathcal{H}}^p$.

Next we apply a poissonization argument. Define

$$
\hat{X}_{i_1} = \delta_{z_1} \left(\hat{P}_n^{(2)} - P_n \right) \dots \left(\hat{P}_n^{(k)} - P_n \right) P_n^{m-k} h \Big|_{z_1 = \hat{Z}_{i_1}^{(1)}},
$$

and

$$
X_{i_1} = \delta_{z_1} \left(\hat{P}_n^{(2)} - P_n \right) \dots \left(\hat{P}_n^{(k)} - P_n \right) P_n^{m-k} h \Big|_{z_1 = Z_{i_1}}.
$$

Let $\hat{\mathbf{Z}} := \left\{ \hat{Z}_{i_2}^{(2)}, \dots, \hat{Z}_{i_k}^{(k)} | i_2, \dots, i_k = 1, \dots, n \right\}$. Note, that conditionally on $\hat{\mathbf{Z}}$, $\hat{X}_{i_1}$ are the bootstrap drops from the sample $\{X_1, \dots, X_n\}$, *and* $E\left[\hat{X}_{i_1} | \hat{\mathbf{Z}} \right] = \frac{1}{n} \sum_{i_1} X_{i_1}$. Apply the symmetrization inequality of Proposition 2.1 in Arcones and Giné (1993), conditionally on $\hat{\mathbf{Z}}$; it gives:

$$
\begin{aligned}
(*) \;&=\; E \left\| n^{-1} \sum\nolimits_{i_1} \left(\hat{X}_{i_1} - E_{|\hat{\mathbf{Z}}} \hat{X}_{i_1} \right) \right\|_{\mathcal{H}}^p \\
&\lesssim\; E \left\| n^{-1} \sum\nolimits_{i_1} \varepsilon_{i_1} \hat{X}_{i_1} \right\|_{\mathcal{H}}^p,
\end{aligned}
$$

where $\{\varepsilon_{i_1}\}$ is a Rademacher sequence independent of all other r.v. in the model. Next by the proof of Lemma 2.1 and Proposition 2.2 of Giné and

Zinn (1990), applied to $\|\cdot\|^p$ rather than $\|\cdot\|$, we obtain:

$$\begin{aligned}
& E\left\|n^{-1}\sum\nolimits_{i_1}\varepsilon_{i_1}\hat{X}_{i_1}\right\|_{\mathcal{H}}^p \\
\lesssim\ & E\left\|n^{-1}\sum\nolimits_{i_1}Q_{i_1}^{(1)}X_{i_1}\right\|_{\mathcal{H}}^p \\
=\ & E\left\|n^{-1}\sum\nolimits_{i_1}Q_{i_1}^{(1)}\delta_{Z_{i_1}}\left(\hat{P}_n^{(2)}-P_n\right)...\left(\hat{P}_n^{(k)}-P_n\right)P_n^{m-k}h\right\|_{\mathcal{H}}^p
\end{aligned}$$

(where the result that we use give the inequality for $Q_{i_1}^{(1)}$ being distributed as a difference of two independent Poisson r.v. (with parameter 1/2). Use the triangle inequality to obtain the inequality for $Q_{i_1}^{(1)}$ being just the Poisson r.v. with parameter 1/2).

Sequential application of this logic to the other arguments (with conditioning on previously introduced Poisson r.v.), and integrating over the distribution of the sample lead to the inequality

$$(*)\lesssim E\left\|n^{-m}\sum_{i_1,...,i_m}Q_{i_1}^{(1)}...Q_{i_k}^{(k)}h\left(Z_{i_1},Z_{i_2},...,Z_{i_m}\right)\right\|_{\mathcal{H}}^p.$$

Note that

$$\begin{aligned}
& E\left\|n^{-m}\sum_{i_1,...,i_m}Q_{i_1}^{(1)}...Q_{i_k}^{(k)}h\left(Z_{i_1},Z_{i_2},...,Z_{i_m}\right)\right\|_{\mathcal{H}}^p \\
\lesssim\ & E\left\|n^{-m}\sum_{i_1,...,i_m}Q_{i_1}^{(1)}...Q_{i_k}^{(k)}...Q_{i_m}^{(m)}h\left(Z_{i_1},Z_{i_2},...,Z_{i_m}\right)\right\|_{\mathcal{H}}^p \\
=\ & E\left\|\frac{1}{n^m}\sum_{i_1,...,i_m}\tilde{h}\left(\tilde{Z}_{i_1},...,\tilde{Z}_{i_m}\right)\right\|_{\tilde{\mathcal{H}}}^p
\end{aligned}$$

by Jensen inequality and the fact that $E\left[Q_{i_{k+1}}^{(k+1)}...Q_{i_m}^{(m)}\right]>0$.

To complete the proof, integrate the bound over the sample measure. ■

Finally, we prove Lemmas 8 and 9.

Proof. (Lemma 8.) (a) For $p=1$, see Corollary 4(i) in Sherman (1994). For $p\geq 2$ use also the Hoffman-Jørgensen inequality, Lemma 15. (b) For $p=1$, see the proof of Corollary 8 in Sherman (1994) (only straightforward notational changes are required). For $p\geq 2$ use also the Hoffman-Jørgensen inequality, Lemma 15.

(c) By Lemma 16 (see the construction of function $\tilde{h}$, $\tilde{z}$, and $\tilde{P}$ there; in particular, $\tilde{h}$ is $\tilde{P}$-degenerate)

$$P\left\|\hat{U}_n^{(m)}h\right\|_{\mathcal{H}}^p \lesssim P\left\|\frac{1}{n^m}\sum_{i_1,\ldots,i_m}\tilde{h}\left(\tilde{Z}_{i_1},\ldots,\tilde{Z}_{i_m}\right)\right\|_{\tilde{\mathcal{H}}}^p.$$

Let $\tilde{U}_n^{(k)}$ denote the U-statistic based on the sample $\left\{\tilde{Z}_{i_1},\ldots,\tilde{Z}_{i_m}\right\}$. Also, for $s \leq m-s$, let ω_s be a permutation, with repetitions, having s elements from the set $\{1,\ldots,s\}$. The permutation $\{1,\ldots,m-s,\omega_s(1),\ldots,\omega_s(s)\}$, therefore, contains $m-s$ distinct elements. Denote by $e_m(\omega_s) = m-s-\#\{\omega_s(1),\ldots,\omega_s(s)\} \geq m-2s$, the number of its non-repeating elements. Denote by $\tilde{h}_{\omega_s}$ the symmetrized version of the function

$$\tilde{h}_{\omega_s}(\tilde{z}_1,\ldots,\tilde{z}_s) = \tilde{h}\left(\tilde{z}_1,\ldots,\tilde{z}_{m-s},\tilde{z}_{\omega_s(1)},\ldots,\tilde{z}_{\omega_s(s)}\right).$$

We can write:

$$\begin{aligned}&\left\|\frac{1}{n^m}\sum_{i_1,\ldots,i_m}\tilde{h}\left(\tilde{Z}_{i_1},\ldots,\tilde{Z}_{i_m}\right)\right\|_{\tilde{\mathcal{H}}}^p\\ \lesssim\ &\left\|\tilde{U}_n^{(m)}\tilde{h}\right\|_{\tilde{\mathcal{H}}}^p + \sum_{1\leq s\leq m/2}\sum_{\omega_s}\left\|n^{-s}\tilde{U}_n^{(m-s)}\tilde{h}_{\omega_s}\right\|_{\tilde{\mathcal{H}}}^p\end{aligned}$$

Note that functions $\tilde{h}_{\omega_s}$ satisfy the condition

$$\tilde{P}^{\#\{\omega_s(1),\ldots,\omega_s(s)\}+1}\tilde{h}_{\omega_s} = 0.$$

Apply operators $\tilde{P}^k$, $k = m-s, m-s+1\ldots,\#\{\omega_s(1),\ldots,\omega_s(s)\}+1$, consecutively to both sides of the Hoeffding decomposition of $\tilde{U}_n^{(m-s)}\tilde{h}_{\omega_s}$ relative to the measure $\tilde{P}$ (the corresponding projections are denoted $\pi_{k,m-s}^{\tilde{P}}$), and conclude that its elements of order $k = 0,1,\ldots,e_m(\omega_s)-1$, are zero:

$$\tilde{U}_n^{(s)}\tilde{h}_{\omega_s} = \sum_{k=e_m(\omega_s)}^{m-s}\binom{m-s}{k}\tilde{U}_n^{(k)}\pi_{k,m-s}^{\tilde{P}}\tilde{h}_{\omega_s}.$$

Next note that every function $\pi_{k,m-s}^{\tilde{P}}\tilde{h}_{\omega_s}$ satisfies the assumptions of part (a) of the Theorem, so that

$$\left\|n^{k/2}\tilde{U}_n^{(k)}\pi_{k,m-s}^{\tilde{P}}\tilde{h}_{\omega_s}\right\|_{\tilde{\mathcal{H}}}^p = O(1)$$

We then have

$$\begin{aligned}&\left\|n^{\frac{m}{2}}\hat{U}_n^{(m)}h\right\|_{\mathcal{H}}^p\\ \lesssim\ &\left\|n^{\frac{m}{2}}\tilde{U}_n^{(m)}\tilde{h}\right\|_{\tilde{\mathcal{H}}}^p+\sum_{1\leq s\leq m/2}\sum_{\omega_s}\sum_{k=e_m(\omega_s)}^{m-s}n^{\frac{m-2s-k}{2}}\left\|n^{k/2}\tilde{U}_n^{(k)}\pi_{k,m-s}^{\tilde{P}}\tilde{h}_{\omega_s}\right\|_{\tilde{\mathcal{H}}}^p.\end{aligned}$$

Next note that in the above sum $m-s-k\leq 0$, and the equality can only be achieved when $k=e_m(\omega_s)=m-2s$, that is when all elements of ω_s are distinct. Finally, note that by the $\tilde{P}$-degeneracy of $\tilde{h}$, for $\omega_s=\{1,2,...,s\}$, $\pi_{m-2s,s}^{\tilde{P}}\tilde{h}_{\omega_s}$ is a constant multiple of the function $\tilde{h}^{[m-2s]}(\tilde{z}_1,...,\tilde{z}_{m-2s})$ (because the other integrals in $\pi_{m-2s,s}^{\tilde{P}}\tilde{h}_{\omega_s}$ involve integrating out non-repeating $\tilde{Z}_i$; also note that $\tilde{h}^{[m-2s]}(\tilde{z}_1,...,\tilde{z}_{m-2s})$ is $\tilde{P}$-degenerate).

Therefore,

$$\begin{aligned}&P\left\|n^{\frac{m}{2}}\hat{U}_n^{(m)}h\right\|_{\mathcal{H}}^p\\ \lesssim\ &P\left\|n^{\frac{m}{2}}\tilde{U}_n^{(m)}\tilde{h}\right\|_{\tilde{\mathcal{H}}}^p+\sum_{1\leq s\leq\frac{m}{2}}P\left\|n^{(m-2s)/2}\tilde{U}_n^{(m-2s)}\tilde{h}^{[m-2s]}\right\|_{\tilde{\mathcal{H}}}^p+O\left(n^{-1/2}\right)\end{aligned}$$

By part (a), the RHS is $O(1)$.

(d) The inequality in the previous display still holds. We check that for $0\leq s\leq\frac{m}{2}$

$$P\left\|n^{(m-2s)/2}\tilde{U}_n^{(m-2s)}\tilde{h}^{[m-2s]}\right\|_{\tilde{\mathcal{H}}}^p=o(1)$$

(where $\tilde{h}^{[m]}=\tilde{h}$). This will follow from part (b) if we show that

$$\left\|\tilde{P}^{m-2s}\left(\tilde{h}^{[m-2s]}\right)^2\right\|_{\tilde{\mathcal{H}}_n}\to 0.$$

This follows from the extra condition in (d) and the construction of $\tilde{h}$ from h. ■

Proof. (Lemma 9.) (a) Define the class of functions

$$\mathcal{H}_n=\left\{h^{\theta,t}=\frac{h_{\theta+n^{-1/2}a_n^{-1}t}-h_\theta}{1+\|t\|^{1/2}}:\|\theta\|,\ n^{-1/2}a_n^{-1}\|t\|\leq\delta_0\right\}.$$

Note that $\mathcal{H}_n$ is Euclidean for the envelope $2H$ because it is a subclass of the Euclidean class

$$\left\{\frac{h_{\theta+t}-h_\theta}{1+\left\|\tilde{t}\right\|^{1/2}}:\|\theta\|,\ \ \|t\|\leq\delta_0,\ \tilde{t}\in\mathbb{R}^d\right\}$$

To prove the lemma, it is enough to show that

$$P\left\{na_n^2\left\|U_n^{(m)}h^{\theta,t}\right\|_{\mathcal{H}_n}>1\right\}=O\left(a_n^{-1}\right). \tag{48}$$

By the Chebyshev inequality,

$$P\left\{na_n^2\left\|U_n^{(m)}h^{\theta,t}\right\|_{\mathcal{H}_n}>1\right\}\leq\left(na_n^2\right)^pP\left\|U_n^{(m)}h^{\theta,t}\right\|_{\mathcal{H}_n}^p.$$

The continuity modulus of class $\mathcal{H}_n$ satisfies

$$\left\|P^m\left(h^{\theta,t}\right)^2\right\|_{\mathcal{H}_n}\leq Cn^{-1/2}a_n^{-1}.$$

By Lemmas 14 and 15,

$$P\left\|U_n^{(m)}h^{\theta,t}\right\|_{\mathcal{H}_n}^p\lesssim\left(n^{-1}\log n\right)^{pm/2}\left(n^{-1/2}a_n^{-1}\right)^{p/2}+\left(n^{-1}\log n\right)^{p(m+1)/2-1}.$$

Therefore, (48) is satisfied if

$$na_n^2\left(n^{-1}\log n\right)^{m/2}\left(n^{-1/2}a_n^{-1}\right)^{1/2}\leq a_n^{-1/p}$$

and

$$na_n^2\left(n^{-1}\log n\right)^{(m+1)/2-1/p}\leq a_n^{-1/p}.$$

These inequalities give

$$a_n\leq\left(n^{m/3-1/2}\left(\log n\right)^{-m/3}\right)^{\frac{1}{1+2/3p}}$$

and

$$a_n\leq\left(n^{\frac{m-1}{4}-\frac{1}{2p}}\left(\log n\right)^{\frac{1}{2p}-\frac{m+1}{4}}\right)^{\frac{1}{1+1/2p}}.$$

from which the result follows immediately.

(b) Let $\mathcal{H}_n$ be as above. Use the inequality obtained in the proof of Lemma 16 (c), rewritten as:

$$\begin{aligned}&P\left\|\hat{U}_n^{(m)}h^{\theta,t}\right\|_{\mathcal{H}}^p\\ \lesssim\;&P\left\|\tilde{U}_n^{(m)}\tilde{h}^{\theta,t}\right\|_{\tilde{\mathcal{H}}}^p+\sum_{1\leq s\leq\frac{m}{2}}P\left\|n^{-s}\tilde{U}_n^{(m-2s)}\tilde{h}^{\theta,t[m-2s]}\right\|_{\tilde{\mathcal{H}}}^p+O\left(n^{-(m+1)/2}\right)\end{aligned}$$

From the additional assumptions made in part (b) of the Lemma, and by construction of functions $\tilde{h}$, we have:

$$\left\| P^m \left(\tilde{h}^{\theta,t}\right)^2 \right\|_{\mathcal{H}_n} \leq Cn^{-1/2} a_n^{-1}, \tag{49}$$

and, for each s, $1 \leq s \leq m/2$,

$$\left\| P^{m-2s} \left(\tilde{h}^{\theta,t[m-2s]}\right)^2 \right\|_{\mathcal{H}_n} \leq Cn^{-1/2} a_n^{-1}. \tag{50}$$

The result now follows from part (a). In particular, notice that we will have

$$\begin{aligned} & P \left\| n^{-s} \tilde{U}_n^{(m-2s)} \tilde{h}^{\theta,t[m-2s]} \right\|_{\tilde{\mathcal{H}}_n}^p \\ \lesssim \;& n^{-sp} \left(n^{-1} \log n\right)^{p(m-2s)/2} \left(n^{-1/2} a_n^{-1}\right)^{p/2} + n^{-sp} \left(n^{-1} \log n\right)^{p(m-2s+1)/2-1} \\ = \;& n^{pm/2} \left(n^{-1} \log n\right)^{p(m-2)/2} \left(n^{-1/2} a_n^{-1}\right)^{p/2} + \left(n^{-1}\right)^{p(m+1)/2-1} (\log n)^{p(m-1)/2-1} \end{aligned}$$

which is dominated by the bound for $P \left\| \tilde{U}_n^{(m)} \tilde{h}^{\theta,t} \right\|_{\tilde{\mathcal{H}}_n}^p$ obtained in part (a) under condition (49). ■

5.4 A Berry-Esséen Bound

Lemma 17 *Let $Z_1, ..., Z_n$ be i.i.d. random variables taking values in a probability space $(\mathcal{Z}, P)$ (P may depend on n), and let $g_\theta : \mathcal{Z}^2 \to \mathbb{R}^d$, $\theta \in \mathbb{R}^d$ be a vector-function, symmetric in z_1, z_2. Let θ_n solve the system of equations*

$$U_n^{(2)} g_\theta^{(l)} = \xi_n^{(l)}(\theta),$$

$l = 1, ..., d$. Assume that there are numbers $\delta_0, K > 0$ such that for all $n \geq 1$:
(i) $P\left\{\sup_{\|\theta\| \leq \delta_0} \|\xi_n(\theta)\| > n^{-1}\right\} \leq Kn^{-1/2}$.
(ii) $\|Pg_0\| \leq Kn^{-1}$.
(iii)$P\|\pi_{1,2}g\|^4 < \infty$, $P\|\pi_{2,2}g\|^2 < \infty$.
(iv) g_θ is twice continuously differentiable in the δ_0-neighborhood of 0, $P-a.e.$, $P\|\partial^2 g_0\|^4 < \infty$, $P\|\partial g_0\|^3 < \infty$, and there is $L(z_1, z_2)$ with $P^2 L^3 < \infty$ such that

$$\left\|\partial^2 g_{\theta_1} - \partial^2 g_{\theta_2}\right\| \leq L \|\theta_1 - \theta_2\|,$$

for all $\|\theta_1\|, \|\theta_2\| \leq \delta_0$.
(v) The $d \times d$ matrix $\Gamma = [\partial P g_0]^{-1} Var(2\pi_{1,2}g_0) [\partial P g_0]^{-1}$ is well defined and

is positive definite (her Var is the variance relative to P).
Then for all $n \geq 1$,

$$\sup_{A \in \mathcal{A}} \left| \int_A dF_{n^{1/2}\theta_n} - \int_A d\Phi_\Gamma \right| \leq c_d n^{-1/2} + c_d P \left\{ \|\theta_n\| > \delta_d \right\},$$

where, for each d, c_d and $\delta_d < \delta_0$ are continuous functions of K, $P \|\pi_{1,2} g_0\|^4$, $P \|\pi_{2,2} g_0\|^2$, $P \|\partial^2 g_0\|^3$, $P \|\partial g_0\|^4$, PL^3.

Proof. (Sketch.) Within the proof, $\lesssim$ denotes an inequality up to a multiplicative constant that may depend on d only, and c_d and δ_d satisfy the conditions of the Theorem (but may change from line to line in the proof). To reduce notation, we assume, without loss of generality, that $\delta_0 \leq 1$ and $K \geq 1$. It is enough to consider the case of $\|\theta_n\| < \delta_0$.

Step 1. Here we prove that, for all $n \geq 1$,

$$P \left\{ \|\theta_n\| > n^{-1/3} \right\} \leq c_d n^{-1/2}.$$

Without loss of generality, assume in this step that $\partial P g_0 = I$ (identity matrix). Use the Taylor expansion around $\theta = 0$ and the Hoeffding decomposition (we omit the index l):

$$\begin{aligned} \xi_n(\theta) &= U_n^{(2)} g_\theta \\ &= U_n^{(2)} g_0 + \left\{ I + P_n \partial \pi_{1,2} g_{\tilde{\theta}} \right\} \theta \end{aligned}$$

where $\tilde{\theta}$ lies between 0 and θ. By our assumptions, the class $\{P_n \partial \pi_{1,2} g_\theta, \ \ \|\theta\| \leq \delta\}$ is Euclidean class for the envelope

$$M(z_1, z_2) = 1 + \|\partial g_0(z_1, z_2)\| + \left\|\partial^2 g_0(z_1, z_2)\right\| + 2\sqrt{d} L(z_1, z_2)$$

(see Lemma (2.13) of Pakes and Pollard (1989) and use the identity, for any f,

$$\partial_l f_\theta = \partial_l f_0 + \int_0^\theta \partial^2_{l,l} f_{\tilde{\theta}} d\theta^{(l)} = \partial_l f_0 + \partial^2_{l,l} f_0 + \int_0^\theta \left(\partial^2_{l,l} f_{\tilde{\theta}} - \partial^2_{l,l} f_0 \right) d\theta^{(l)}.$$

The envelope is made bigger than necessary to simplify notation later on). Therefore, (by the fact that $P \partial \pi_{1,2} g_\theta = 0$ for each non-random θ, because $P \pi_{1,2} g_\theta = 0$ and $P \|\partial g_0\| < \infty$; then apply the bound for the suprema for the second moment to deal with the randomness in $\tilde{\theta}$)

$$P \left\{ \|P_n \partial \pi_{1,2} g_{\tilde{\theta}}\| \geq \frac{1}{2} \right\} \lesssim n^{-1} P M^2.$$

Then

$$
\begin{array}{rl}
& P\left\{\|\theta_n\|>n^{-1/3}\right\} \\
\lesssim & P\left\{4\left\|U_n^{(2)}g_0\right\|>n^{-1/3}\right\}+P\left\{2\left\|\xi_n\left(\theta\right)\right\|>n^{-1/3}\right\}+n^{-1}PM^2 \\
\lesssim & Kn^{-1/2},
\end{array}
$$

by the Hoeffding decomposition, Maximal and Chebyshev inequalities.

Step 2. Obtain the representation: for all $n\geq 1$,

$$
\theta^{(l_1)}=U_n^{(2)}g_*^{(l_1)}+C_{l_1,l_2,l_3}\theta^{(l_1)}\theta^{(l_2)}+\zeta_n^{(l_1)} \tag{51}
$$

where C_{l_1,l_2,l_3} are constants, function $g_*^{(l_1)}(z_1,z_2)$ does not depend on θ and satisfies $P^2g_*=0$, $Var\left[2\pi_{1,2}g_*\right]=\Gamma$, $P^2\left\|\pi_{1,2}g_*\right\|^3<\infty$, $P^2\left\|\pi_{2,2}g_*\right\|^2<\infty$; and

$$
P\left\{\|\zeta_n\|>c_dn^{-1}\right\}\leq n^{-1/2}c_d+P\left\{\|\theta_n\|>\delta_dn^{-1/3}\right\}.
$$

In this step, assume, without loss of generality, that $\partial Pg_0=I$. By the Hoeffding decomposition and the Taylor expansion, for each component l_1,

$$
\begin{aligned}
0 &= U_n^{(2)}g_\theta^{(l_1)}-\xi_{n,\theta}^{(l_1)} \\
&= U_n^{(2)}\left(g_0^{(l_1)}-P^2g_0^{(l_1)}\right)+\left\{\left(I+B_n\right)\theta\right\}^{(l_1)}-C_{l_1,l_2,l_3}\theta^{(l_2)}\theta^{(l_3)} \\
&\quad +\tilde{\zeta}_n^{(l_1)}\left(\theta\right)
\end{aligned} \tag{52}
$$

where $\left(B_n\right)_{l_1,l_2}=2P_n\partial_{l_2}\pi_{1,2}g_0^{(l_1)}$, $C_{l_1,l_2,l_3}=-\frac{1}{2}\partial_{l_2,l_3}P^2g_0^{(l_1)}$, and

$$
\begin{aligned}
\tilde{\zeta}_n^{(l_1)}\left(\theta\right) &= \frac{1}{2}\left(\partial_{l_2,l_3}P^2g_{\tilde{\theta}}^{(l_1)}-\partial_{l_2,l_3}P^2g_0^{(l_1)}\right)\theta^{(l_2)}\theta^{(l_3)} \\
&\quad +P_n\partial_{l_2,l_3}\pi_{1,2}g_{\tilde{\theta}}^{(l_1)}\theta^{(l_2)}\theta^{(l_3)} \\
&\quad +U_n^{(2)}\partial_{l_2}\pi_{2,2}g_{\tilde{\theta}}^{(l_1)}\theta^{(l_2)}-\xi_{n,\theta}^{(l_1)}+P^2g_0^{(l_1)}
\end{aligned}
$$

and $\tilde{\theta}$ is in between 0 and θ (in fact, $\tilde{\theta}$ is different for each l_1 and each term above, but we will ignore this distinction). Using the identity

$$
\begin{aligned}
\left(I+B_n\right)^{-1} &= I-B_n\left(I+B_n\right)^{-1} \\
&= I-B_n+B_n^2\left(I+B_n\right)^{-1}
\end{aligned}
$$

we can rewrite (52) as

$$
\theta^{(l_1)}=U_n^{(2)}g_*^{(l_1)}+C_{l_1,l_2,l_3}\theta^{(l_1)}\theta^{(l_2)}+\zeta_n^{(l_1)}\left(\theta\right)
$$

where

$$g_*^{(l_1)}(z_1,z_2) = -\left(g_0^{(l_1)}(z_1,z_2) - P^2 g_0^{(l_1)}\right) - 2\partial_{l_2}\pi_{1,2}g_0^{(l_1)}(z_1)\cdot\pi_{1,2}g_0^{(l_2)}(z_2),$$

(in particular, the random vector g_* satisfies the above properties), and

$$\begin{aligned}\zeta_n^{(l_1)}(\theta) &= U_n^{(2)}\left\{-(I-B_n)\left(g_0 - P^2 g_0\right) - g_*\right\}^{(l_1)} \\ &\quad - \left\{B_n^2 (I+B_n)^{-1} U_n^{(2)}\left(g_0 - P^2 g_0\right)\right\}^{(l_1)} \\ &\quad + \left[B_n (I+B_n)^{-1}\right]_{l_1,l_2} C_{l_2,l_3,l_4}\theta^{(l_3)}\theta^{(l_4)} \\ &\quad - \left\{(I+B_n)^{-1}\tilde{\zeta}_n(\theta)\right\}^{(l_1)}.\end{aligned}$$

Using the bound in Step 1, one can see that all terms in this expression satisfy the restriction on $\zeta_n(\theta)$.

Step 3. Now use representation (51) to obtain the Berry-Esséen bound.

Consider the system of equations

$$\theta^{(l_1)} = \gamma^{(l_1)} + C_{l_1 l_2 l_3}\theta^{(l_2)}\theta^{(l_3)}. \tag{53}$$

By the Implicit Function Theorem and the Taylor expansion, there are numbers $\delta^* > 0$, $K_1 > 0$, and $b_{l_1 l_2 l_3}$, continuously depending on $C_{l_1 l_2 l_3}$, such that if $\|\gamma\| \le \delta^*$, and θ is the solution of (53) satisfying $\|\theta\| \le \delta^*$, then

$$\theta^{(l_1)} = \gamma^{(l_1)} + b^{(2)}_{l_1 l_2 l_3}\gamma^{(l_2)}\gamma^{(l_3)} + \phi^{(l_1)}(\gamma)$$

and

$$\|\phi(\gamma)\| \le K_1 \|\gamma\|^3.$$

Let $\gamma_n = U_n^{(2)} g_* + \zeta_n$. By the Hoeffding decomposition, the properties of g_* and ζ_n, and the Chebyshev inequality,

$$P\left\{\|\gamma_n\| > \delta^*\right\} \lesssim c_d n^{-1/2} + P\left\{\|\theta\| > \delta_d n^{-1/3}\right\}, \tag{54}$$

and

$$\begin{aligned} & P\left\{\|\phi(\gamma_n)\| \ge n^{-1}, \|\gamma_n\| \le \delta^*\right\} \\ \lesssim\ & n^{-1/2}\left(P\|\pi_{1,2}g_*\|^3 + P\|\pi_{2,2}g_*\|^2\right).\end{aligned}$$

Next, consider the statistic T_n defined as follows:

$$T_n^{(l_1)} = U_n^{(2)} g_* + b_{l_1 l_2 l_3}\frac{1}{n^2}\sum_{i\ne j}\left\{\pi_{1,2}g_{*i}^{(l_2)}\cdot\pi_{1,2}g_{*j}^{(l_3)}\right\},$$

where $\pi_{1,2}g_{*i}^{(l_2)} \equiv \pi_{1,2}g_*^{(l_2)}(Z_i)$.

Note that

$$\begin{aligned} & n^{1/2}P\left\{\left|\theta_n^{(l_1)} - T_n^{(l_1)}\right| \geq n^{-1}, \|\gamma_n\| \leq \delta^*, \|\theta_n\| \leq \delta^*\right\} \\ \lesssim \; & P\left\|\pi_{1,2}g_*\right\|^2 + P\left\|\pi_{1,2}g_*\right\|^3 + P\left\|\pi_{2,2}g_*\right\|^2. \end{aligned} \tag{55}$$

Using (54) and (55), we conclude that

$$\theta_n = T_n + \xi_n$$

$$P\left\{\|\xi_n\| > n^{-1}\right\} \leq c_d n^{-1/2} + P\left\{\|\theta\| > \delta_d n^{-1/3}\right\}. \tag{56}$$

T_n has a form of a U-statistic of order 2 with zero mean. Its variance is $n^{-1}\Gamma$ up to a term of order $O(n^{-2})$.

The Berry-Esséen bound for T_n follows from Theorem 2 of Bolthausen and Götze (1993):

$$\sup_{A\in\mathcal{A}}\left|\int_A dF_{n^{1/2}T_n} - \int_A d\Phi_\Gamma\right| \leq c_d n^{-1/2}.$$

The conclusion of the theorem follows from the last result and (56) by an argument similar to that at the end of Section 5.2.3. ■