Multiplier Bootstrap for Empirical Processes

Yu-Chin Hsu†

Institute of Economics

Academia Sinica

First version: November 11, 2015

This version: September 4, 2016

† E-mail: ychsu@econ.sinica.edu.tw. 128 Academia Road, Section 2, Nankang, Taipei, 115 Taiwan. Tel: 886-2-27822791 ext. 322. Fax: 886-2-2785-3946.

Acknowledgement: I thank Hiroaki Kaido, Robert P. Lieli and Xiaoxia Shi for helpful suggestions and comments. All errors are my responsibility. I gratefully acknowledges the research support from Ministry of Science and Technology of Taiwan (MOST103-2628-H-001-001-MY4) and Career Development Award of Academia Sinica, Taiwan.

Abstract

Multiplier bootstrap (MB) has been used to approximate the limiting processes of empirical processes in various papers. In this paper, we consider multiplier bootstrap in three cases. First, we consider MB for standard empirical process. Second, we extend the MB to account for estimation effects of the pre-estimated parameters or unknown nonparametric functions. Last, we consider MB for Nadaraya-Waston nonparametric kernel estimators.

JEL classification: C01,C15

Keywords: Empirical Process, Multiplier Bootstrap, Nonparametric Estimator.

1 Introduction

The multiplier bootstrap (MB), based on the multiplier central limit theorem, in Section 2.9 of van der Vaart and Wellner (1996, VW hereafter) has been applied in several papers to approximate Gaussian processes so as to obtain valid critical values when the limiting null distribution of a test statistic involves empirical processes. For example, Barrett and Donald (2003), Donald, Hsu and Barrett (2012), and Donald and Hsu (2013, 2014) use MB to construct critical values for tests of stochastic dominance relations. Hsu (2015) applies MB to construct tests for conditional treatment effects. Hsu, Lai and Lieli (2015) and Hsu, Kan and Lai (2015) use it to construct confidence bands for local quantile treatment effects and structural quantile treatment effects. Hsu, Liu and Shi (2015) apply it to obtain critical values for tests for generalized regression monotonicity. Bugni, Canay and Shi (2015) use it to construct a specification test for moment inequality models. Kaido and Santo (2014) employ the MB to estimate the asymptotic distribution of set estimators.

MB is an alternative to the nonparametric bootstrap or the parametric bootstrap. To approximate the Gaussian limiting process, the idea of MB is first to multiply the estimated influence functions with mean zero and unit variance pseudo-random variables that are independent of the sample path and then to sum up these items. An advantage of MB over conventional bootstrap is that we can avoid recomputing the estimator in each bootstrap repetition, which can be time-consuming when there are several steps to obtain the estimators. One disadvantage of MB is that one needs to obtain an analytic expression for the influence function representation of the estimators.

In this paper, we consider MB in three cases. In the first case, we consider the standard empirical process and the MB method. In this case, the only unknown parameters to be estimated in the estimated influence functions is the unconditional mean of each index. In the second case, we consider the MB method to approximate the empirical processes generated from multiple-step estimation. In this case, we need to consider the estimation effects of the pre-estimated parameters or unknown functions. In the last case, we consider the MB method for empirical processes from Nadaraya-Waston nonparametric kernel estimators. The main difference between our MB method and most of the papers in the literature is that we allow the data generating processes to vary with the sample size. This is important if one is interested in deriving the local power of a test or if one wants to obtain the uniform size of tests involving moment inequalities.

The idea of MB is the same as the "score bootstrap" proposed by Kline and Santos (2012). The main difference between our paper and theirs is that we focus on cases in which the empirical process contains infinitely many elements and they focus on a finite number of elements. Our paper is also related to the multiplier bootstrap of Chernozhukov, Chetverikov and Kato (2013). The main difference is that they focus on the approximation of a maximum of a sum of the Gaussian random vectors but we focus on the approximation of an empirical process. We also extend our results to account for estimation effects.

The rest of this paper is organized as follows. In Section 2, we derive the results for the standard empirical process and show the validity of the MB method. In Section 3, we extend the results to account for estimation effects. We consider empirical process and the MB method for Nadaraya-Waston nonparametric kernel estimators in Section 4. Section 5 concludes and all mathematical proofs are deferred to the Appendix.

2 Empirical Processes and the Multiplier Bootstrap

We define some notation. Let $\{W_{ni}: \ 1 \le i \le n, \ n \ge 1\}$ denotes a row-wise i.i.d. triangular array of $\mathcal{W}$-valued random variables where $\mathcal{W}$ is a Borel subset of R^{d_w}. Also, let P_n denotes the distribution function of W_{ni}. Let E_P denote the expectation under the distribution P; we will simply write E when there is no confusion. Let $\mathcal{Q} = \{q(w,t): \ t \in \mathcal{T}\}$ be a collection of scalar-valued functions with index set $\mathcal{T}$. Let $\widehat{\Phi}_n(t) = n^{-1/2}\sum_{i=1}^n \big(q(W_{ni},t) - E[q(W_{ni},t)]\big)$ denote the empirical process which is an indexed collection of random variables, $\{\widehat{\Phi}_n(t): \ t \in \mathcal{T}\}$. Under suitable conditions, we show that $\widehat{\Phi}_n(t)$ converges to a Gaussian process. We call $\{(q(W_{ni},t) - E[q(W_{ni},t)])/\sqrt{n}: \quad t \in \mathcal{T}\}$ the influence functions of the $\widehat{\Phi}_n(t)$; that is, the measure of the dependence of the estimator on the value of one of the points in the sample. We call $\widehat{\Phi}_n(t)$ the asymptotically linear representation of the limiting process.

Let $\{U_1, U_2, ...\}$ be a sequence of i.i.d. pseudo random variables with $E[U] = 0$ and $E[U^2] = 1$ that are independent of the sequence $\{W_{ni}: \ 1 \le i \le n, \ n \ge 1\}$. Define the simulated process as

$$\widehat{\Phi}_n^u(t) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n U_i \cdot \Big(q(W_{ni},t) - \bar{q}_n(t)\Big),$$

where $\bar{q}_n(t) = n^{-1}\sum_{i=1}^n q(W_{ni},t)$. We call methods based on the simulated process $\widehat{\Phi}_n^u(\tau)$ as the "Multiplier Bootstrap (MB)" in the paper. The idea of MB is to multiply the estimated influence functions, $(q(W_{ni},t) - \bar{q}_n(t))/\sqrt{n}$, by the pseudo random random variables U_i's and sum up these terms to approximate the limiting process. We call $(q(W_{ni},t) - \bar{q}_n(t))/\sqrt{n}$ estimated influence function as $\bar{q}_n(t)$ has to be estimated in advance. If we replace $\bar{q}_n(t)$ with $E[q(W_{ni},t)]$, then $n^{-1/2}\sum_{i=1}^n U_i \cdot \Big(q(W_{ni},t) - E[q(W_{ni},t)]\Big)$ is the simulated process defined in VW Section 2.9.

Let Q be an envelope function for $\mathcal{Q}$, i.e., Q is a real function on $\mathcal{W}$ such that $|q(w,t)| \le Q(w)$ for all $t \in \mathcal{T}$, for all $w \in \mathcal{W}$. Let $h_2(\cdot,\cdot)$ be a covariance kernel on $\mathcal{T} \times \mathcal{T}$. Let $\mathcal{H}_2$ be the collection of all possible covariance kernel function on $\mathcal{T} \times \mathcal{T}$. For any pair of $h_2^{(1)}$ and $h_2^{(2)}$, we define the distance between them as

$$d(h_2^{(1)}, h_2^{(2)}) = \sup_{t_1,t_2 \in \mathcal{T}} |h_2^{(1)}(t_1,t_2) - h_2^{(2)}(t_1,t_2)|. \tag{2.1}$$

Let $\Rightarrow$ denote weak convergence of a stochastic process and the weak convergence is in the sense of VW.

We introduce some assumptions.

Assumption 2.1 *Assume that*

(i) $\{W_{ni} : 1 \le i \le n, n \ge 1\}$ *is a row-wise i.i.d. triangular array with* $W_{n,i} \sim P_n$.

(ii) $\big|q(w,t)\big| \le Q(w)$ *for all* $w \in \mathcal{W}$, *for all* $t \in \mathcal{T}$ *for some envelope function* $Q(w)$.

(iii) $\{q(W_{ni}, t) : \; t \in \mathcal{T}, \; 1 \le i \le n, \; n \ge 1\}$ *is manageable w.r.t. the envelope functions* $\{Q(W_{ni}) : i \le n, 1 \le n\}$ *in the sense of Definition 7.9 of Pollard (1990).*

(iv) $E_{P_n}[Q^{2+\delta}(W_{ni})] \le C < \infty$ *for all* P_n *for some* $\delta > 0$.

(v) There exists $h_2 \in \mathcal{H}_2$ *such that* $d(h_{2,P_n}, h_2) \to 0$ *where* $h_{2,P}$ *is defined as*

$$h_{2,P}(t_1, t_2) = E_P\big[\big(q(W, t_1) - E_P[q(W, t_1)]\big) \cdot \big(q(W, t_2) - E_P[q(W, t_2)]\big)\big]$$

Assumption 2.2 *Assume that*

(i) $\{U_i : \; i = 1, \ldots, \}$ *is a sequence of i.i.d. random variables with probability distribution* $\mathbb{P}_u$ *and* $E[U] = 0$, $E[U^2] = 1$, *and* $E[|U|^{2+\delta}] < M$ *for some* $M > 0$, *and where* δ *is as in Assumption 2.1.*

(ii) $\mathcal{U}$ *is independent of* $\{W_{ni} : \; 1 \le i \le n, \; n \ge 1\}$.

We summarize the first set of main results in the following theorem and we prove it in Appendix.

Theorem 2.1 *Suppose that Assumption 2.1 holds. Then* $\widehat{\Phi}_n(\cdot) \Rightarrow \Phi_{h_2}(\cdot)$ *where* $\Phi_{h_2^*}(\cdot)$ *denote a mean zero Gaussian process with covariance kernel* $h_2^* \in \mathcal{H}_2$. *Furthermore, if Assumption 2.2 also holds, then* $\widehat{\Phi}^u_n(\cdot) \Rightarrow \Phi_{h_2}(\cdot)$ *conditional on the sample path* $\{W_{ni} : \; 1 \le i \le n, \; n \ge 1\}$ *with probability 1 which is denoted as* $\widehat{\Phi}^u_n(\cdot) \overset{p}{\Rightarrow} \Phi_{h_2}(\cdot)$.[1]

Remarks:

1. The first part of Theorem 2.1 is not new in the literature, but to the best of our knowledge, the second part is new.
2. Assumption 2.1 (iii) holds automatically if $\mathcal{T}$ contains only a finite number of elements. Assumption 2.1 (v) holds automatically if $P_n = P$ for all n.
3. Assumption 2.1 provides sufficient condition for the Functional Central Limit Theorem (FCLT) of Pollard (1990, Theorem 10.6). Theorem 2.1 is less general than FCLT of Pollard (1990), because FCLT of Pollard (1990) applies to nonparametric estimation as in Donald, Hsu and Barrett (2012), and Andrews and Shi (2015), but Theorem 2.1 does not. The main reason is that Assumption 2.1 does not allow f functions to vary with n and in nonparametric estimation case, the f functions include the bandwidth which depends on n. We will consider such extension in Section 4.

[1] The conditional weak convergence is in the sense of Section 2.9 of VW and Chapter 2 of Kosorok (2008). To be more specific, $\Psi^u_n \overset{p}{\Rightarrow} \Psi$ in the metric space $(\mathbb{D}, d)$ if and only if $\sup_{f \in BL_1} |E_u f(\Psi^u_n) - Ef(\Psi)| \overset{p}{\to} 0$ and $E_u f(\Psi^u_n)^* - E_u f(\Psi^u_n)_* \overset{p}{\to} 0$, where the subscript u in E_u indicates conditional expectation over the weights U_i's given the remaining data, BL_1 is the space of functions $f : \mathbb{D} \to R$ with Lipschitz norm bounded by 1, and $f(\Psi^u_n)^*$ and $f(\Psi^u_n)_*$ denote measurable majorants and minorants with respect to the joint data including the U_i's. We use the notation $\Psi^u_n \overset{a.s.}{\Rightarrow} \Psi$ to mean the same thing except with all $\overset{p}{\to}$'s replaced by $\overset{a.s.}{\to}$'s. Note that by Lemma 1.9.2 (ii) of VW, it is true that $\Psi^u_N \overset{p}{\Rightarrow} \Psi$ if and only if every subsequence k_N of N has a further subsequence ℓ_N of k_N such that $\Psi^u_{\ell_N} \overset{a.s.}{\Rightarrow} \Psi$.

4. Andrews (1994) shows that sets of functions that satisfy Pollard's entropy condition are manageable, i.e., Assumption 2.1 (iii) would hold. Pollard's entropy condition is given in (4.2) of Andrews (1994). Andrews (1994) also gives lots of examples satisfying Pollard's entropy condition. These examples include Vapnik-Vervonenkis (VC) classes, Lipschitz functions indexed by finite dimensional parameters and infinite dimensional classes of smooth functions. Theorem 2 and Theorem 3 of Andrews (1994) show how we can generate more classes of functions that satisfy Pollard's entropy condition from sets of functions that are known to satisfy Pollard's entropy condition. Please see Andrews (1994) for more details. Therefore, our Theorem 2.1 can be applied to a lots of cases.

5. The second part implies that the simulated empirical process can approximate the limiting process well. This is a generalization of conditional multiplier central limit theorem of VW in that we allow the distribution P_n to vary and we replace the $E_{P_n}[q(W_n, t)]$ with its sample analog $\bar{q}_n(\tau)$. Note that if $P_n = P$ for all n, we can weaken the conditions in Assumptions 2.1 and 2.2 such that $\delta = 2$. Theorem 2.1 is important especially when we consider local alternatives in which the P_n varies.

6. If the U_i's are standard normals, then for each fixed t,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i \cdot \Big(q(W_{ni}, t) - \bar{q}_n(t) \Big) \overset{d}{\sim} \mathcal{N}(0, \hat{V}_t), \text{ where}$$
$$\hat{V}_t = \frac{1}{n} \sum_{i=1}^{n} \Big(q(W_{ni}, t) - \bar{q}_n(t) \Big)^2$$

which is the traditional asymptotic approximation.

7. Theorem 2.1 can be extended to cases where $q(w, t)$ are vector-valued by using the Cramer-Wold device as in the proof of Lemma A1(a) of Andrews and Shi (2013).

3 Accounting for Estimation Effects

In this section, we extend Theorem 2.1 to allow for the case in which in first step, we need to estimate some parametric components or nonparametric components before we define the empirical processes. In other words, the influence functions may include some parametric or nonparametric components to be estimated in addition to $E[q(W_{ni}, t)]$. Therefore, to obtain the estimated influence functions for the MB, we would need to estimate these additional element. For example, in Donald and Hsu (2014), $\widehat{\Phi}_n(\cdot) = \sqrt{n}(\widehat{F}_j(\cdot) - F_j(\cdot))$, where $F_1(\cdot)$ is the distribution function (CDF) of the potential outcome $Y(1)$ in a treatment effect model. In this case, $q_n(W, t) = (D \cdot 1(Y \leq t))/p(X) - \big(F_1(t|X)/p(X)\big) \cdot \big(D - p(X)\big) - F_1(t)$ where $-\infty < y_\ell \leq t \leq y_u < \infty$, $W = (Y, D, X')$, Y is the outcome observed, D is the treatment status and X is a set of covariates so that the unconfoundedness assumption would hold conditional on X, $p(X) = P(D = 1|X)$ is the propensity score and $F_1(\cdot|X)$ is the conditional CDF of $Y(1)$ conditional on X. In this example, $q_n(W, t)$ includes $p(X)$ and $F_1(t|X)$ that we need to estimate in advance. Please see Example 3.1 below for more details.

Let $\widehat{\Phi}_n(\cdot)$ denote an empirical process that has an asymptotic linear representation such that $\widehat{\Phi}_n(t) = n^{-1/2}\sum_{i=1}^n q_n(W_{ni}, t) + o_P(1)$ for $t \in \mathcal{T}$ where $o_p(1)$ holds uniformly over $t \in \mathcal{T}$. With loss of generality, we assume that $E_{P_n} q_n(W_{ni}, t) = 0$ and because of this, we have q_n depending on n as well. Let $\hat{q}_n(W_{ni}, t)$ be the estimated influence functions and define the simulated process $\Phi_n^u(t)$, as $\Phi_n^u(t) = n^{-1/2}\sum_{i=1}^n U_i \cdot \hat{q}_n(W_{ni}, t)$. We first make the following high-level assumptions and later, we consider two examples in which we give low level sufficient conditions.

Assumption 3.1 *Assume that*

(i) $\{W_{ni} : 1 \le i \le n, n \ge 1\}$ *is a row-wise i.i.d. triangular array with* $W_{n,i} \sim P_n$.

(ii) $|q_n(w,t)| \le Q_n(w)$ *for all* $w \in \mathcal{W}$, *for all* $t \in \mathcal{T}$ *for some envelope function* $Q_n(w)$.

(iii) $\{q_n(W_{ni}, t) : \ t \in \mathcal{T}, \ 1 \le i \le n, \ n \ge 1\}$ *is manageable w.r.t. the envelope functions* $\{Q_n(W_{ni}) : i \le n, 1 \le n\}$ *in the sense of Definition 7.9 of Pollard (1990).*

(iv) $E_{P_n}[q_n(W_{ni}, t)] = 0$ *for all* $t \in \mathcal{T}$ *and* $E_{P_n}[Q_n^{2+\delta}(W_{ni})] \le C < \infty$ *for all* P_n *for some* $\delta > 0$.

(v) There exists $h_2 \in \mathcal{H}_2$ *such that* $d(h_{2,P_n}, h_2) \to 0$ *where* $h_{2,P}$ *is defined as* $h_{2,P}(t_1, t_2) = E_P[q(W, t_1) \cdot q(W, t_2)]$.

Let $\{\hat{Q}_n(W_{ni}) : i \le n, 1 \le n\}$ denote the envelope functions so that $|\hat{q}_n(w,t)| \le \hat{Q}_n(w)$ for all $w \in \mathcal{W}$ and for all $t \in \mathcal{T}$. Note that here the envelope functions depend on some unknown functions that need to be estimated as well. This will not cause any problem in the proof because to show the simulated process results, all the arguments are conditional on the sample path. As a result, these estimated components are treated as constant functions. Please see Example 3.2 for details. We define $\rho_n^u(t_1, t_2) = \sum_{i=1}^n \hat{q}_n(W_{ni}, t_1) \cdot \hat{q}_n(W_{ni}, t_2)$ for all $t_1, t_2 \in \mathcal{T}$.

Assumption 3.2 *Assume that*

(i) $\{\hat{q}_n(W_{ni}, t) : \ t \in \mathcal{T}, \ 1 \le i \le n, \ n \ge 1\}$ *is manageable w.r.t. the envelope functions* $\{\hat{Q}_n(W_{ni}) : 1 \le i \le n, 1 \le n\}$ *in the sense of Definition 7.9 of Pollard (1990).*

(ii) $d(\rho_n^u(t_1, t_2), h_2) \xrightarrow{p} 0$.

(iii) There exists $0 < \delta_1 < \delta$ *such that*

$$\frac{1}{n}\sum_{i=1}^n \hat{Q}_n^2(W_{ni}) - \frac{1}{n}\sum_{i=1}^n Q_n^2(W_{ni}) \xrightarrow{p} 0, \text{ and}$$
$$\frac{1}{n}\sum_{i=1}^n \hat{Q}_n^{2+\delta_1}(W_{ni}) - \frac{1}{n}\sum_{i=1}^n Q_n^{2+\delta_1}(W_{ni}) \xrightarrow{p} 0.$$

Assumption 3.1 is the same as Assumption 2.1 except that the functions q and Q functions depend on n and that $E_{P_n}[q_n(W_{ni}, t)] = 0$. Assumption 3.2 is a high-level assumption and below we give two examples in which Assumption 3.2 holds under low-level conditions. Assumption 3.2 is sufficient for us to apply the proof for Theorem 2.1 to show the following theorem.

Theorem 3.1 *Suppose that Assumption 3.1 holds. Then* $\widehat{\Phi}_n(\cdot) \Rightarrow \Phi_{h_2}(\cdot)$. *Furthermore, if Assumptions 2.2 and 3.2 also hold, then* $\widehat{\Phi}^u_n(\cdot) \overset{p}{\Rightarrow} \Phi_{h_2}(\cdot)$.

We consider two examples. One is the CDF estimator of the potential outcomes of Donald and Hsu (2014) that is briefly discussed above and the other one is the empirical process used in Hsu (2015) to test hypotheses about conditional average treatment effects.

Example 3.1 As we discussed above, Donald and Hsu (2014) propose an inverse probability weighted estimator for the CDF of the potential outcome $Y(1)$, $F_1(y)$, in a treatment effect model.[2] Under the unconfoundness assumption $Y(1)\perp D\big|X$, $F_1(y)$ is identified by

$$F_1(y) = E\Big[\frac{D \cdot 1(Y \le y)}{p(X)}\Big].$$

The proposed estimator is defined as

$$\widehat{F}_1(y) = n^{-1}\sum_{i=1}^{n}\frac{D_i \cdot 1(Y_i \le y)}{\hat{p}(X_i)}\Big/ n^{-1}\sum_{i=1}^{n}\frac{D_i}{\hat{p}(X_i)},$$

where $y \in [y_l, y_u]$ for some $-\infty < y_l < y_u < \infty$ and $\hat{p}(x)$ is the series logit estimator for the propensity score function used in Hirano, Imbens and Ridder (2003). Under regularity conditions, Donald and Hsu (2014) show that

$$\sqrt{n}(\widehat{F}_1(y) - F_1(y)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} q(W_i, y) + o_p(1) \text{ with}$$
$$q(W_i, y) = \frac{D_i \cdot 1(Y_i \le y)}{p(X_i)} - \frac{F_1(y|X_i)}{p(X_i)} \cdot \big(D_i - p(X_i)\big) - F_1(y),$$

where $o_p(1)$ terms are uniform over $y \in [y_l, y_u]$ and $W = (Y, D, X')$.[3] Donald and Hsu (2014) show that $\{F_1(y|X) : \ y \in [y_l, y_u]\}$ which are weakly monotonic functions in y satisfies the Pollard's entropy conditions with envelope functions being 1's. $\{1(Y \le y) : \ y \in [y_l, y_u]\}$ is a VC class, so it is manageable with respect to envelope functions of 1's. By Theorem 3 of Andrews (1994) and Lemma E1 of Andrews and Shi (2013) and given that the propensity score function is assumed to be bounded away from zero, we have that $\{q(W_i, y) : \ y \in [y_l, y_u], \ 1 \le i \le n, \ n \ge 1\}$ is manageable with respect to the constant envelope functions M for some $M > 0$. Donald and Hsu (2014) show that $\Phi_n(y) = \sqrt{n}(\widehat{F}_1(y) - F_1(y)) \Rightarrow \Phi_{h_2}(y)$ where $h_2(y_1, y_2) = E[q(W_i, y_1)q(W_i, y_2)]$.

Let $\hat{p}(x)$ and $\widehat{F}_1(y|x)$ be the estimators for $p(x)$ and $F_1(y|x)$, respectively, and the MB simulated process is defined as

$$\Phi^u_n(y) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i \cdot \hat{q}(W_i, y) + o_p(1),$$
$$\hat{q}(W_i, y) = \frac{D_i \cdot 1(Y_i \le y)}{\hat{p}(X_i)} - \frac{\widehat{F}_1(y|X_i)}{\hat{p}(X_i)} \cdot \big(D_i - \hat{p}(X_i)\big) - \widehat{F}_1(y)$$

[2]We focus on the $Y(1)$ case, and the $Y(0)$ case is similar.

[3]Donald and Hsu (2014) focus on the case in which $P_n = P$ and with suitable modification, we can extend their results to allow P_n to drift as in Hsu (2015).

in which $\widehat{F}_1(y|x)$ is monotonically increasing in y and bounded between 0 and 1 for all x. Also, $\hat{p}(x)$ and $\widehat{F}_1(y|x)$ that are uniformly consistent in that

$$\sup_{x\in\mathcal{X}} |\hat{p}(x) - p(x)| \xrightarrow{p} 0, \text{ and } \sup_{x\in\mathcal{X},\ y\in[y_l,y_u]} |\widehat{F}_1(y|x) - F_1(y|x)| \xrightarrow{p} 0. \tag{3.1}$$

Let $\underline{\hat{p}}_n = \inf_{x\in\mathcal{X}} \hat{p}(X_i)$ and let $\hat{Q}(W_{ni}) = 2\underline{\hat{p}}_n + 1$. Then it is straightforward to see that Assumption 3.2(i) would hold. Note that (3.1) is sufficient to show that3.2(ii) would hold. (3.1) also implies that (iii) of Assumption 3.2 holds. Therefore, they can show that $\Phi_n^u(y) \overset{p}{\Rightarrow} \Phi_{h_2}(y)$. Based on these results, Donald and Hsu (2014) propose tests for the stochastic dominance relations between the potential outcomes and please see Donald and Hsu (2014) for details.

Example 3.2 Hsu (2015) proposes tests for the null hypothesis that conditional average treatment effects (CATE) is nonnegative for every covariate value.[4] Let $\mu_0(x) = E_P[Y(0)|X = x]$ and $\mu_1(x) = E_P[Y(1)|X = x]$, and $CATE(x)$ is defined as $\text{CATE}(x) = \mu_1(x) - \mu_0(x)$. Let X be a d_x-dimensional vector of covariates with $d_x \geq 1$ and X has a compact support $\mathcal{X}$, say $\times[0,1]^{d_x}$. The null hypothesis is defined as

$$H_0 : CATE(x) \geq 0, \quad \text{for all } x \in \mathcal{X}. \tag{3.2}$$

Hsu (2015) uses Andrews and Shi (2013) instrumental variable function approach to transform the conditional moment inequality into infinitely many unconditional ones without information loss. The set of the instrumental variable functions is the set of the indicator functions of countable hyper cubes:

$$\begin{aligned} &\mathcal{G} = \{g_\ell(\cdot) = 1(\cdot \in C_\ell) : \ell \equiv (x, r) \in \mathcal{L}\}, \text{ where} \\ &C_\ell = \Big(\times_{j=1}^{d_x}[x_j, x_j + r]\Big) \cap \mathcal{X} \text{ and} \\ &\mathcal{L} = \Big\{(x, q^{-1}) : \ q \cdot x \in \{0, 1, 2, \cdots, q-1\}^{d_x}, \text{ and } q = 1, 2, \cdots\Big\}, \end{aligned} \tag{3.3}$$

such that for each q, $\{C_\ell : \ell \in \mathcal{L} \text{ and } r = q^{-1}\}$ forms a partition of $\mathcal{X}$. Under unconfoundedness assumption, the null hypotheses in (3.2) is equivalent to

$$H_0 : \nu(\ell) \equiv E_P\Big[-g_\ell(X)\Big(\frac{DY}{p(X)} - \frac{(1-D)Y}{1-p(X)}\Big)\Big] \leq 0, \text{ for all } \ell \in \mathcal{L}. \tag{3.4}$$

Let the estimator for $\nu(\ell)$ be

$$\hat{\nu}(\ell) = \frac{1}{n}\sum_{i=1}^{n} -g_\ell(X_i)\Big(\frac{D_iY_i}{\hat{p}(X_i)} - \frac{(1-D_i)Y_i}{1-\hat{p}(X_i)}\Big).$$

Under regularity conditions, it is shown that

$$\sup_{\ell\in\mathcal{L}} \Big|\sqrt{n}\big(\hat{\nu}(\ell) - \nu(\ell)\big) - \frac{1}{\sqrt{n}}\sum_{i=1}^{N} q(W_i, \ell)\Big| = o_{\mathcal{P}}(1),$$

[4]For notational simplicity, here we consider CATE(X) such that the unconfoundedness assumption would hold when we condition on X. Hsu (2015) considers CATE(X_1) in which X_1 is a strict subset of X and conditioning on X_1 alone, the unconfoundedness assumption will not hold.

where $W \equiv \{Y, D, X\}$ and

$$q(W,\ell) = g_\ell(X)\Big(\frac{(1-D)Y}{1-p(X)} - \frac{DY}{p(X)} + (D-p(X))\Big(\frac{\mu_0(X)}{1-p(X)} + \frac{\mu_1(X)}{p(X)}\Big)\Big) - \nu(\ell).$$

Note that $\{g_\ell(X):\ \ell \in \mathcal{L}\}$ is a VC class of functions, so it is manageable with respect to envelope functions of 1's. By Theorem 3 of Andrews (1994) and Lemma E1 of Andrews and Shi (2013), $\{q(W_i,\ell):\ \ell \in \mathcal{L},\ 1 \le i \le n,\ n \ge 1\}$ is manageable with respect to envelope functions $\{m(W_i) + E[m(W_i)]:\ 1 \le i \le n,\ n \ge 1\}$ and

$$m(W) = \Big|\frac{(1-D)Y}{1-p(X)} - \frac{DY}{p(X)} + (D-p(X))\Big(\frac{\mu_0(X)}{1-p(X)} + \frac{\mu_1(X)}{p(X)}\Big)\Big|.$$

Define $h_{2,P}(\ell_1,\ell_2) = E_P[q(W,\ell_1)q(W,\ell_2)] = Cov_P(q(W,\ell_1)q(W,\ell_2))$. If $\lim_{n\to\infty} d(h_{2,P_n}, h_2) = 0$ for some $h_2 \in \mathcal{H}$, Lemma 3.1 of Hsu (2015) shows that $\sqrt{n}(\hat{\nu}_n(\cdot) - \nu_n(\cdot)) \Rightarrow \Psi_{h_2}(\cdot)$. For all $\ell \in \mathcal{L}$, define the simulated stochastic processes $\Psi^u(\ell)$ as

$$\Psi^u(\ell) = \frac{1}{\sqrt{n}}\sum_{i=1}^n U_i \cdot \hat{q}(W_i,\ell), \text{ and} \tag{3.5}$$

$$\hat{q}(W_i,\ell) = g_\ell(X_i)\Big(\frac{(1-D_i)Y_i}{1-\hat{p}(X_i)} - \frac{D_iY_i}{\hat{p}(X_i)} + (D_i-\hat{p}(X_i))\Big(\frac{\hat{\mu}_0(X_i)}{1-\hat{p}(X_i)} + \frac{\hat{\mu}_1(X_i)}{\hat{p}(X_i)}\Big)\Big) - \hat{\nu}(\ell), \tag{3.6}$$

where $\hat{\mu}_0(x)$ and $\hat{\mu}_1(x)$ are the series estimators for $\mu_0(x)$ and $\mu_1(x)$ such that $\sup_{x\in\mathcal{X}} |\hat{\mu}_0(x) - \mu_0(x)| = o(1)$ and $\sup_{x\in\mathcal{X}} |\hat{\mu}_1(x) - \mu_1(x)| = o(1)$. Let

$$\hat{Q}(W_i) = \hat{M}(W_i) + \hat{\overline{M}},$$
$$\hat{M}(W_i) = \Big|\Big(\frac{(1-D_i)Y_i}{1-\hat{p}(X_i)} - \frac{D_iY_i}{\hat{p}(X_i)} + (D_i-\hat{p}(X_i))\Big(\frac{\hat{\mu}_0(X_i)}{1-\hat{p}(X_i)} + \frac{\hat{\mu}_1(X_i)}{\hat{p}(X_i)}\Big)\Big)\Big|,$$
$$\hat{\overline{M}} = \frac{1}{n}\sum_{i=1}^n \hat{M}(W_i).$$

Then one can show that Assumption 3.2 holds. Then Lemma 4.1 of Hsu (2015) shows that $\Psi^u_n(\ell) \overset{p}{\Rightarrow} \Psi_{h_2}(\ell)$. These results are used to derive the uniformity of the proposed tests and please see Hsu (2015) for details.

4 Nadaraya-Waston Nonparametric Kernel Estimator

This section we discuss the MB method to approximate the empirical processes obtained from Nadaraya-Waston nonparametric kernel estimators such as Donald, Hsu and Barrett (2012). We introduce more notations. Let Z denote the conditioning variable with support $\mathcal{Z}$ which is a Borel subset of R^{d_z} and let $f_{P,z}(z)$ denote the probability density function of z on $\mathcal{Z}$ under P. Note that Z can overlap, non-overlap or be a subvector of W.[5] Let $\mathcal{Z}_o$ denote a collection of interior points of $\mathcal{Z}$. Let $\mathcal{Z}_o^\epsilon = \{z' :\ |z' - z|_\infty \le \epsilon$ for some $z \in \mathcal{Z}_o\}$ for some $\epsilon > 0$ not depending on n and P where $|\cdot|_\infty$ is the supremum norm, and we assume that $\mathcal{Z}_o^\epsilon \subseteq \mathcal{Z}$. Let $\mu_P(t,z) = E_P[q(W,t)|Z=z]$ denote the conditional mean of $g(W,t)$ conditioning on $Z = z$ under P. For notational simplicity, we suppress the subscripts P of $f_{P,z}(z)$ and

[5]Of course, we exclude the case where $Z = W$.

$\mu_P(t,z)$ whenever there is no confusion. Let $K(\cdot): R^{d_z} \to R$ denote the kernel function, and h the bandwidth depending on sample size. Let $\hat{\mu}_n(t,z)$ denote the Nadaraya-Waston nonparametric kernel estimator for $\mu(t,z)$ that is defined as

$$\hat{\mu}_n(t,z) = \Big(\frac{1}{nh^{d_z}}\sum_{i=1}^n K\Big(\frac{Z_i - z}{h}\Big)q(W_i,t)\Big)\Big/\Big(\frac{1}{nh^{d_z}}\sum_{i=1}^n K\Big(\frac{Z_i-z}{h}\Big)\Big).$$

The empirical process of our interest is $\widehat{\Phi}_n(\cdot) = \sqrt{nh^{d_z}}(\hat{\mu}_n(\cdot,z) - \mu(\cdot,z))$, and under suitable conditions, we will show that

$$\widehat{\Phi}_n(\cdot) = \frac{1}{\sqrt{nh^{d_z}}}\sum_{i=1}^n \frac{1}{f_z(z)}K\Big(\frac{Z_i-z}{h}\Big)\cdot(q(W_i,t) - \mu(\cdot,z)) + o_p(1). \tag{4.1}$$

where the $o_p(1)$ result holds uniformly over $t \in \mathcal{T}$.[6] Based on (4.1), we defined the BP simulated processes as

$$\widehat{\Phi}^u_n(\cdot) = \frac{1}{\sqrt{nh^{d_z}}}\sum_{i=1}^n U_i \cdot \frac{1}{\hat{f}_z(z)}K\Big(\frac{Z_i-z}{h}\Big)\cdot(q(W_i,t) - \hat{\mu}_n(\cdot,z)), \tag{4.2}$$

where $\hat{f}_{z,n}(z)$ is the Nadaraya-Waston nonparametric kernel estimator for $f_z(z)$, i.e., $\hat{f}_{z,n}(z) = (nh^{d_z})^{-1}\sum_{i=1}^n K\big((Z_i - z)/h\big)$. We make the following assumptions. Also, define $\zeta(t_1,t_2,z) = E_P[q(W,t_1)q(W,t_2)|Z=z]$.

Assumption 4.1 *Assume that*

(i) $\{(W_{ni}, Z_{ni}) : 1 \le i \le n, n \ge 1\}$ *is a row-wise i.i.d. triangular array with* $(W_{ni}, Z_{ni}) \sim P_n$.

(ii) $\big|q(w,t)\big| \le Q(w)$ *for all* $w \in \mathcal{W}$, *for all* $t \in \mathcal{T}$ *for some envelope function* $Q(w)$.

(iii) $\{q(W_{ni},t):\ t \in \mathcal{T},\ 1 \le i \le n,\ n \ge 1\}$ *is manageable w.r.t. the envelope functions* $\{Q(W_{ni}) : i \le n, 1 \le n\}$ *in the sense of Definition 7.9 of Pollard (1990).*

(iv) $E_{P_n}[Q^4(W_n)|Z_n = z] \le C < \infty$ *for all* $z \in \mathcal{Z}_o^\epsilon$ *for all* P_n *for some* $\delta > 0$.

(v) $\lim_{n\to\infty} f_{P_n,z}(z) = f$ *for some* $f > 0$.

(vi) $\lim_{n\to\infty}\sup_{t_1,t_2\in\mathcal{T}}|\zeta_n(t_1,t_2,z) - \zeta(t_1,t_2,z)|$ *for some* $\zeta(t_1,t_2,z)$.

(vii) There exists $h_2 \in \mathcal{H}_2$ *such that* $d(h_{2,P_n,z}, h_2) \to 0$ *where* $h_{2,P}$ *is defined as*

$$\begin{aligned} h_{2,P_n,z}(t_1,t_2)\|K\|_2^2 \cdot f^{-1}_{P_n,z}(z)E\big[(q(W_{ni},t_1) - \mu_n(t_1,z))\cdot(q(W_{ni},t_2) - \mu_n(t_2,z))\big] \\ = \|K\|_2^2 \cdot f^{-1}_{P_n,z}(z)\cdot\big(\zeta_n(t_1,t_2,z) - \mu_n(t_1,z)\mu_n(t_2,z)\big) \end{aligned}$$

and $\|K\|_2 = (\int K^2(u)du)^{1/2}$.

Assumption 4.2 *Assume that*

(i) $f_{P_n,z}(z) \ge \delta_z$ *for some* $\delta_z > 0$ *and for all* $z \in \mathcal{Z}_o^\epsilon$ *and for all* n.

[6] We could have put z_n instead of z to allow the conditioning value varying with n and P_n, but we only consider z here for notational simplicity. The proof for the case in which z_n is allowed is identical to the case in which $z_n = z$ for all n.

(ii) $f_{P_n,z}(z)$ *is twice continuously differentiable,* $|f_{P_n,z}(z)| \leq M$*,* $|f'_{P_n,z}(z)| \leq M$ *and* $|f''_{P_n,z}(z)| \leq M$ *for some* $0 < M < \infty$ *on* $\mathcal{Z}_o^\epsilon$ *for all* n*.*

(iii) $\mu_{P_n}(t,z)$ *is twice continuously differentiable,* $\sup_{t\in\mathcal{T}} |\mu_{P_n}(t,z)| \leq M$*,* $\sup_{t\in\mathcal{T}} |\mu'_{P_n}(t,z)| \leq M$ *and* $\sup_{t\in\mathcal{T}} |\mu''_{P_n}(t,z)| \leq M$ *on* $\mathcal{Z}_o^\epsilon$ *for all* n*.*

(iv) $\zeta(t_1,t_2,z)$ *is twice continuously differentiable,* $\sup_{t_1,t_2\in\mathcal{T}} |\zeta(t_1,t_2,z)| \leq M$*,* $\sup_{t_1,t_2\in\mathcal{T}} |\zeta'(t_1,t_2,z)| \leq M$*, and* $\sup_{t_1,t_2\in\mathcal{T}} |\zeta''(t_1,t_2,z)| \leq M$ *on* $\mathcal{Z}_o^\epsilon$ *for all* n*.*

Assumption 4.3 *Assume that*

(i) The $K(\cdot)$ *is a non-negative symmetric bounded kernel with a compact support in* R^{d_z} *(say* $\times_{j=1}^{d_z}[-1,1]$*).*

(ii) $\int K(u)du = 1$ *and* $\int u_j K(u)du = 0$ *for* $j = 1,\ldots,d_z$*.*

(iii) $h \to 0$*,* $nh^{d_z} \to \infty$ *and* $nh^{d_z+4} \to 0$ *as* $n \to \infty$*.*

Theorem 4.1 *Suppose that Assumptions 4.1-4.3 hold. Then* $\widehat{\Phi}_n(\cdot) \Rightarrow \Phi_{h_2}(\cdot)$*. Furthermore, if Assumption 2.2 also holds, then* $\widehat{\Phi}_n^u(\cdot) \overset{p}{\Rightarrow} \Phi_{h_2}(\cdot)$*.*

Remarks:

1. Assumption 4.1 is similar to Assumption 2.1 and 3.1, but it is adjusted to account for nonparametric estimation.
2. Assumption 4.2 is the smoothness conditions we need for nonparametric estimations that are standard in the literature.
3. Assumption 4.3 imposes conditions on kernel function and bandwidth. Assumption 4.3(iii) requires undersmoothing so the bias term is asymptotically negligible. This is standard if one wants the nonparametric estimator to be asymptotically normally distributed with mean zero.
4. We can extend Theorem 4.1 to the local polynomial kernel estimators.

Example 4.3 Donald, Hsu and Barrett (2012) propose tests for stochastic dominance relations between conditional CDFs under different model setups. To do this, they need to estimate the conditional CDFs. In Section 3.4, they consider Nadaraya-Waston nonparametric kernel estimator for $F(y|x)$ denoted as $\widehat{F}(y|z)$

$$\widehat{F}(y|z) = \Big(\frac{1}{nh^{d_z}}\sum_{i=1}^n K\Big(\frac{Z_i - z}{h}\Big)\cdot 1(Y_i \leq y)\Big)\Big/\Big(\frac{1}{nh^{d_z}}\sum_{i=1}^n K\Big(\frac{Z_i - z}{h}\Big)\Big).$$

Note that $\{1(Y \leq y) : y \in [y_\ell, y_u]\}$ is a VC class of functions, so Assumption 4.1(iii) holds with envelope functions $Q(w) = 1$. Hence, Assumption 4.1(iv) holds. They consider the case in which $P_n = P$, so Assumption 4.1(v) and (iv) hold automatically. They assume that $F(y|z)$ are twice continuously differentiable on the compact support $[y_\ell, y_u]$ and this is sufficient for Assumption 4.2(iii) and (iv).

Under regularity conditions, they show that

$$\begin{aligned}&\widehat{\Phi}_n(\cdot) = \sqrt{nh^{d_z}}(\widehat{F}(y|z) - F(y|z))\\=&\frac{1}{\sqrt{nh^{d_z}}}\sum_{i=1}^{n}\frac{1}{f_z(z)}K\Big(\frac{Z_i - z}{h}\Big)\cdot(1(Y_i \leq y) - F(y|z)) + o_p(1)\end{aligned}$$

where the $o_p(1)$ result holds uniformly over $y \in [y_\ell, y_u]$. In turn, they show in their Theorem 3.7 that $\widehat{\Phi}_n(\cdot) \Rightarrow \Phi_{h_2}(y)$ with $h_2(y_1, y_2) = f_z^{-1}(z)\|K\|_2^2(F(y_1|z) - F(y_1|z)F(y_2|z))$ for $y_1 \leq y_2$ by showing that

$$\widehat{\Phi}_n(\cdot) = \frac{1}{\sqrt{nh^{d_z}}}\sum_{i=1}^{n}\frac{1}{f_z(z)}K\Big(\frac{Z_i - z}{h}\Big)\cdot(1(Y_i \leq y) - F(y|z)) + o_p(1)$$

where the $o_p(1)$ result holds uniformly over y. The BP simulated processes as is

$$\widehat{\Phi}_n^u(\cdot) = \frac{1}{\sqrt{nh^{d_z}}}\sum_{i=1}^{n}U_i\cdot\frac{1}{\hat{f}_z(z)}K\Big(\frac{Z_i - z}{h}\Big)\cdot(1(Y_i \leq z) - \widehat{F}(y|z)), \tag{4.3}$$

where $\hat{f}_{z,n}(z)$ is the Nadaraya-Waston nonparametric kernel estimator for $f_z(z)$. In Theorem 3.8, they show $\widehat{\Phi}_n^u(\cdot) \overset{p}{\Rightarrow} \Phi_{h_2}(y)$.

5 Conclusion

In this paper, we consider sufficient conditions for the validity of MB for empirical processes in three cases: (1) we consider standard empirical process; (2) we extend the MB to account for estimation effects of the pre-estimated parameters or unknown nonparametric functions; (3) we consider MB for Nadaraya-Waston nonparametric kernel estimators. For cases (2) and (3), we first provide high-level conditions and the give examples with low-level conditions.

APPENDIX

We first restate Theorem 10.6 (FCLT) of Pollard (1990). Recall $(\Omega, \mathcal{F}, \mathbb{P})$ is the underlying probability space and We introduce additional notations. Let the triangular array be $\{f_{ni}(\omega, t): \ \tau \in \mathcal{T}, \ 1 \leq i \leq n, \ n = 1, 2, \ldots\}$ with $\{f_{ni}\}$ are independent within each row where $\omega \in \Omega$. Let $\{F_{ni}: \ 1 \leq i \leq n, \ n = 1, 2, \ldots\}$ denote the envelopes functions in that $|f_{ni}(\omega, t)| \leq F_{ni}(\omega)$ for all $1 \leq i \leq n$ and for all $\omega \in \Omega$. Define

$$X_n(\omega, t) = \sum_{i=1}^{n} (f_{ni}(\omega, t) - E f_{ni}(\cdot, t)),$$

$$\rho_n(t_1, t_2) = \Big[\sum_{i=1}^{n} E\Big(f_{ni}(\cdot, t_1) - f_{ni}(\cdot, t_2)\Big)^2\Big]^{1/2}.$$

Also, assume that $\rho_n(t_1, t_2) \to \rho(t_1, t_2)$ for all $t_1, t_2 \in \mathcal{T}$. We restate Theorem 10.6 (FCLT) of Pollard (1990).

Functional Central Limit Theorem of Pollard (1990): *Suppose that the processes from the triangular array $\{f_{ni}(\omega, t)\}$ are independent within rows and satisfy:*

(i) the $\{f_{ni}\}$ are manageable with respect to $\{F_{ni}(\omega)\}$ in the sense of Definition 7.9 of Pollard (1990),

(ii) $H(t_1, t_2) = \lim_{n\to\infty} E[X_n(\cdot, t_1) X_n(\cdot, t_2)]$ exists for every $t_1, t_2 \in \mathcal{T}$,

(iii) $\limsup \sum_{i=1}^{n} E[F_{ni}^2(\cdot)] < \infty$,

(iv) $\sum_{i=1}^{n} E[F_{ni}^2(\cdot) \cdot 1(F_{ni} > \epsilon)] \to 0$ for each $\epsilon > 0$,

(v) $\rho(t_1, t_2) = \lim_{n\to\infty} \rho_n(t_1, t_2)$ for all $t_1, t_2 \in \mathcal{T}$, and for all deterministic sequence $\{t_{1n}\}$ and $\{t_{2n}\}$ if $\rho(t_{1n}, t_{2n}) \to 0$, then $\rho_n(t_{1n}, t_{2n}) \to 0$.

Then $X_n(\cdot) \Rightarrow \Phi_H(\cdot)$. □

A Proof for Main Results

Proof of Theorem 2.1: The first part of proof is done by checking that conditions (i)-(v) of FCLT are satisfied in our case.

Define $f_{ni}(\omega, t) = \big(g(W_{ni}(\omega), t) - E[g(W_{ni}(\cdot), t)]\big)/\sqrt{n}$ and $F_{ni}(\omega) = \big(Q(W_{ni}(\omega)) + E[Q(W_{ni}(\cdot), t)]\big)/\sqrt{n}$. Then $X_n(t) = \widehat{\Phi}_n(t)$. Also, given that we assume that W_{ni} are i.i.d. within rows, so we have $\rho_n(t_1, t_2) = E[\big(g(W_{ni}(\omega), t_1) - E[g(W_{ni}(\cdot), t_1)]\big) \cdot (g(W_{ni}(\omega), t_2) - E[g(W_{ni}(\cdot), t_2)])]$. Also, $H(t_1, t_2) = \lim_{n\to\infty} \rho_n(t_1, t_2)$ and $\lim_{n\to\infty} h_{2,P_n}(t_1, t_2) = h_2(t_1, t_2)$ by Assumption 2.1 (v).

Assumption 2.1(iii) and Lemma E1 of Andrews and Shi (2013) imply (i) of FCLT. Given that $H(t_1, t_2) = \lim_{n\to\infty} \rho_n(t_1, t_2) = \lim_{n\to\infty} h_{2,P_n}(t_1, t_2) = h_2(t_1, t_2)$ by Assumption 2.1 (v), this is sufficient for (ii) of FCLT. Note that for any random variable $W \geq 0$, we have $E[W^2] \leq 1 + E[W^\delta]$ for any $\delta \geq 2$. Therefore, Assumption 2.1(iv) implies that $\limsup \sum_{i=1}^{n} E[F_{ni}^2] = \limsup E[Q(W_{ni})^2] \leq$

$\limsup(1 + E[Q^\delta(W_{ni})]) \leq 1 + C < \infty$. Therefore, (iii) of FCLT holds. Note that

$$\sum_{i=1}^{n} E[F_{ni}^2(\cdot) \cdot 1(F_{ni} > \epsilon)] = nE[F_{ni}^2(\cdot) \cdot 1(F_{ni} > \epsilon)] \leq nE\Big[\frac{F_{ni}^{2+\delta}}{\epsilon^\delta}\Big]$$
$$= nE_{P_n}\Big[\Big(\frac{Q(W_{ni}) + E_{P_n}(Q(W_{ni}))}{\sqrt{n}}\Big)^{2+\delta}\Big]$$
$$= n^{-\delta}\epsilon^{-\delta}E_{P_n}\big[\big(Q(W_{ni}) + E_{P_n}Q(W_{ni})\big)^{2+\delta}\big] \leq n^{-\delta}\epsilon^{-\delta}2E_{P_n}Q^{2+\delta}(W_{ni}) \to 0,$$

where the first equality holds because the data are i.i.d., the first inequality holds because $1(F_{ni} > \epsilon) \leq (F_{ni}/\epsilon)^\delta$ for any $\delta > 0$, the second and third equalities hold by definition, the last inequality holds because $\big(Q(W_{ni}) + E_{P_n}Q(W_{ni})\big)^{2+\delta} \leq \big(Q(W_{ni})^{2+\delta} + E_{P_n}Q(W_{ni})\big)^{2+\delta}$ and $\big|E_{P_n}Q(W_{ni})\big|^{2+\delta} \leq E_{P_n}Q^{2+\delta}(W_{ni})$, and given that $E_{P_n}Q^{2+\delta}(W_{ni}) \leq C$ for all n, so the last result holds. This implies that (v) of FCLT holds. Last, Assumption 2.1 (v) is sufficient for (v) of FCLT. Therefore, conditions (i)-(v) of FCLT hold in our case, so we have $X_n(\cdot) \Rightarrow \Phi_H(\cdot)$ or equivalently, $\widehat{\Phi}_n(\cdot) \Rightarrow \Phi_{h_2}(\cdot)$.

To show the second part, for each $\omega \in \Omega$, let $f_{ni}^u(\omega, t) = U_i \cdot (q(W_{ni}(\omega), t) - \bar{g}_n(\omega, t))/\sqrt{n}$. Let the triangular array be $\{f_{ni}^u(\omega, t) :\ t \in \mathcal{T},\ 1 \leq i \leq n,\ n \geq 1\}$ and let $X_n^u(\omega, t) = \sum_{i=1}^n f_{ni}^u(\omega, t)$. It is obvious that $X_n^u(\omega, t) = \phi_n^u(\omega, t)$. Let E_u denote the expectation with respect to U_i's. Let $\rho_n^u(t_1, t_2)(\omega) = \sum_{i=1}^n E_u f_{ni}^u(\omega, t_1) f_{ni}^u(\omega, t_2) = n^{-1}\sum_{i=1}^n (q(W_{ni}(\omega), t_1) - \bar{g}_n(\omega, t_1))(q(W_{ni}(\omega), t_2) - \bar{g}_n(\omega, t_2))$. Let $H_n^u(t_1, t_2)(\omega) = E_u X_n^u(\omega, t_1) X_n^u(\omega, t_2)$ and it is obvious that $H_n^u(t_1, t_2) = \rho_n^u(t_1, t_2)(\omega)$.

We first show that $d(\rho_n^u(t_1, t_2), h_2) \xrightarrow{p} 0$. Note that $\rho_n^u(t_1, t_2) = n^{-1}\sum_{i=1}^n (q(W_{ni}, t_1) - \bar{q}_n(t_1))(q(W_{ni}, t_2) - \bar{q}_n(t_2)) = n^{-1}\sum_{i=1}^n q(W_{ni}, t_1) q(W_{ni}, t_2) - \bar{g}_n(t_1)\bar{g}_n(t_2)$. Consider the triangular $\{q(W_{ni}, t) :\ 1 \leq i \leq n\}$ that is manageable with respect to the envelope functions $\{Q(W_{ni}) :\ 1 \leq i \leq n\}$. By Assumption 2.1, we also have $n^{-1}\sum_{i=1}^n E_{P_n}Q^{2+\delta}(W_{ni}) = E_{P_n}Q^{2+\delta}(W_{ni}) < \infty$. Then by Lemma E2 of Andrews and Shi (2013b), we have

$$\sup_{t \in \mathcal{T}} \big|\bar{q}_n(W_{ni}, t) - E_{P_n} q(W_{ni}, t)\big| \xrightarrow{p} 0.$$

Then by Slutsky's theorem, we have

$$\sup_{t_1, t_2 \in \mathcal{T}} \big|\bar{q}_n(W_{ni}, t_1)\bar{q}_n(W_{ni}, t_2) - E_{P_n} q(W_{ni}, t_1) E_{P_n} q(W_{ni}, t_2)\big| \xrightarrow{p} 0.$$

By the same argument, we have

$$\sup_{t_1, t_2 \in \mathcal{T}} \big|n^{-1}\sum_{i=1}^{n} q(W_{ni}, t_1) q(W_{ni}, t_2) - E_{P_n} q(W_{ni}, t_1) \cdot q(W_{ni}, t_2)\big| \xrightarrow{p} 0.$$

These imply that

$$\sup_{t_1, t_2 \in \mathcal{T}} \big|n^{-1}\sum_{i=1}^{n} q(W_{ni}, t_1) q(W_{ni}, t_2) - \bar{q}_n(W_{ni}, t_1)\bar{q}_n(W_{ni}, t_2) - E_{P_n} q(W_{ni}, t_1) \cdot q(W_{ni}, t_2)$$
$$+ E_{P_n} q(W_{ni}, t_1) E_{P_n} q(W_{ni}, t_2)\big| \xrightarrow{p} 0.$$

Or equivalently, $d(\rho_n^u(t_1, t_2), h_{2,P_n}) \xrightarrow{p} 0$. By Assumption 2.1, $d(h_{2,P_n}, h_2) \to 0$, and by triangular inequality, we have $d(\rho_n^u(t_1, t_2), h_2) \leq d(\rho_n^u(t_1, t_2), h_{2,P_n}) + d(h_{2,P_n}, h_2) \xrightarrow{p} 0$. From Footnote XXX, we show the second part by showing that that for any subsequence $\{b_n\}$ of $\{n\}$, there exists a further subsequence of $\{k_n\}$ such that $\widehat{\Phi}^u_{k_n} \Rightarrow \Psi_{h_2}$ conditional on sample path almost surely.

Note that by similar argument, we have $n^{-1}\sum_{i=1}^{n}(Q(W_{ni})+\bar{Q}_n)^2 - E_{P_n}(Q(W_{ni})+E_{P_n}Q(W_{ni}))^2 \xrightarrow{p} 0$ and for some $0<\delta_1<\delta$, $n^{-1}\sum_{i=1}^{n}(Q(W_{ni})+\bar{Q}_n)^{\delta_1} - E_{P_n}(Q(W_{ni})+E_{P_n}Q(W_{ni}))^{2+\delta_1} \xrightarrow{p} 0$. Also, by Assumption 2.1, there exists $0<M<\infty$ such that $E_{P_n}(Q(W_{ni})+E_{P_n}Q(W_{ni}))^2 < M$ and $E_{P_n}(Q(W_{ni})+E_{P_n}Q(W_{ni}))^{\delta_1} < M$ for all n. Therefore, there exists $0<M<\infty$ such that for any subsequence $\{b_n\}$ of $\{n\}$, there exists a further subsequence $\{k_n\}$ of $\{b_n\}$ such that

$$
\begin{aligned}
&d(\rho^u_{k_n}(t_1,t_2),h_2) \xrightarrow{a.s.} 0,\\
&\limsup k_n^{-1}\sum_{i=1}^{k_n}(Q(W_{k_n i})+\bar{Q}_{k_n})^2 \le M \text{ almost surely, and}\\
&\limsup k_n^{-1}\sum_{i=1}^{k_n}(Q(W_{k_n i})+\bar{Q}_{k_n})^{2+\delta_1} \le M \text{ almost surely.}
\end{aligned}
$$

Define

$$
\begin{aligned}
\Omega_1 \equiv \Big\{\omega\in\Omega : \; & d(\rho^u_{k_n}(t_1,t_2),h_2)(\omega)\to 0,\\
&\limsup k_n^{-1}\sum_{i=1}^{k_n}(Q(W_{k_n i})+\bar{Q}_{k_n})^2(\omega) \le M \textit{ and}\\
&\limsup k_n^{-1}\sum_{i=1}^{k_n}(Q(W_{k_n i})+\bar{Q}_{k_n})^{2+\delta_1}(\omega) \le M.\Big\}
\end{aligned}
$$

It is obvious that $P(\Omega_1)=1$. We show that conditioning on any given $\omega\in\Omega_1$, the conditions (i)-(v) of FCLT hold. For (i), by Assumption 2.1, we have $\{q(W_{ni}(\omega),t):\ t\in\mathcal{T},\ 1\le i\le k_n,\ \ k_n\ge 1\}$ is manageable with respect to the envelope functions $\{Q(W_{ni}(\omega),t): 1\le i\le k_n,\ k_n\ge 1\}$. Then by Lemma E1(b) of Andrews and Shi (2013), we have $\{(q(W_{ni},t_1)-\bar{g}_n(t_1)):\ t\in\mathcal{T},\ 1\le i\le k_n,\ k_n\ge 1\}$ is manageable with respect to the envelope functions $\{Q(W_{ni}(\omega),t): 1\le i\le k_n,\ k_n\ge 1\}$. The by Lemma E1(a) of Andrews and Shi (2013), we have $\{U_i\cdot(q(W_{k_n i}(\omega),t)-\bar{g}_{k_n}(\omega,t))/\sqrt{k_n}:\ t\in\mathcal{T}\ 1\le i\le k_n,\ \ k_n\ge 1\}$ is manageable with respect to the envelope functions $\{|U_i|\cdot(Q(W_{k_n i}(\omega))+\bar{Q}_{k_n}(\omega))/\sqrt{k_n}:\ 1\le i\le k_n,\ \ k_n\ge 1\}$. Let $F^u_{k_n i}=|U_i|\cdot(Q(W_{k_n i}(\omega))+\bar{Q}_{k_n}(\omega))/\sqrt{k_n}$. Then, $\{f^u_{k_n i}(\omega,t):\ t\in\mathcal{T},\ 1\le i\le n,\ k_n\ge 1\}$ is manageable with respect to the envelope functions $\{F^u_{k_n i}(\omega):\ 1\le i\le k_n,\ k_n\ge 1\}$, so (i) of FCLT holds. Recall that $E_u X^u_{k_n}(\omega,t_1)X^u_{k_n}(\omega,t_2)=\rho^u_{k_n}(t_1,t_2)(\omega)=k_n^{-1}\sum_{i=1}^{k_n}(q(W_{k_n i}(\omega),t_1)-\bar{g}_{k_n}(\omega,t_1))(q(W_{k_n i}(\omega),t_2)-\bar{g}_{k_n}(\omega,t_2))\to h_2(t_1,t_2)$, so (ii) holds. Note that $E_u\sum_{i=1}^{k_n}E_u(F^u_{k_n i}(\omega))^2 = k_n^{-1}\sum_{i=1}^{k_n}(Q(W_{k_n i}(\omega))+\bar{Q}_{k_n}(\omega))^2\le M<\infty$, so (iii) holds. Next, by the similar argument in the first part,

$$
\begin{aligned}
&\sum_{i=1}^{k_n}E_u(F^u_{k_n i}(\omega))^2\cdot(F^u_{k_n i}(\omega)>\epsilon)\le \epsilon^{-\delta_1}\sum_{i=1}^{k_n}E_u(F^u_{k_n i}(\omega))^{2+\delta_1}\\
=&n^{-\delta_1}\epsilon^{-\delta_1}n^{-1}\sum_{i=1}^{k_n}E_u|U_i|^{2+\delta_1}(Q(W_{k_n i}(\omega))+\bar{Q}_{k_n}(\omega))^{2+\delta_1}\\
\le&M\cdot k_n^{-\delta_1}\epsilon^{-\delta_1}k_n^{-1}\sum_{i=1}^{k_n}(Q(W_{k_n i}(\omega))+\bar{Q}_{k_n}(\omega))^{2+\delta_1}\to 0,
\end{aligned}
$$

where the last inequality holds because by Assumption 2.2, $E[|U|^{2+\delta}] \leq C$, there exists $0 < M < \infty$ such that $E[|U|^{2+\delta_1}] \leq M$ when $0 < \delta_1 < \delta$. Hence, (iv) holds. Last, (v) holds because $d(\rho^u_{k_n}(t_1, t_2)(\omega), h_2) \to 0$. By FCLT, $\widehat{\Phi}^u_{k_n}(\omega) \Rightarrow \Psi_{h_2}$ for all $\omega \in \Omega_1$ or equivalently $\widehat{\Phi}^u_{k_n}(\omega) \overset{a.s.}{\Rightarrow} \Psi_{h_2}$. Therefore, for any sequence $\{b_n\}$ of $\{n\}$, there exists a further subsequence $\{k_n\}$ of $\{b_n\}$ such that $\widehat{\Phi}^u_{k_n}(\omega) \overset{a.s.}{\Rightarrow} \Psi_{h_2}$, so we have $\widehat{\Phi}^u_n(\omega) \overset{p}{\Rightarrow} \Psi_{h_2}$ and the second part holds. □

Proof of Theorem 3.1: The proof is identical to that of Theorem 2.1, so we omit it. □

Proof of Theorem 4.1: In this proof, we assume that $d_z = 1$ for notational simplicity and the proof for $d_z \geq 2$ is identical. We introduce more notations. Let $r(t,z) = \int_{\mathcal{W}} q(w,t) f(w,z) dz$ where $f(w,z)$ is the joint probability density of W and Z, and it is obvious that $r(t,z) = \mu(t,z) \cdot f_z(z)$. Let

$$\hat{r}_n(t,z) = \frac{1}{nh} \sum_{i=1}^{n} K\big(\frac{Z_i - z}{h}\big) q(W_i, t),$$

and it is obvious that $\hat{r}(t,z) = \hat{\mu}_n(t,z) \cdot \hat{f}_{z,n}(z)$. We first derive the process result of $\sqrt{nh}(\hat{r}_n(t,z) - r(t,z))$. Define

$$f_{ni}(t) = \frac{1}{\sqrt{nh}} K\big(\frac{Z_i - z}{h}\big) q(W_i, t), \quad F_{ni} = \frac{1}{\sqrt{nh}} K\big(\frac{Z_i - z}{h}\big) Q(W_i, t).$$

We have

$$\begin{aligned} &\sqrt{nh}(\hat{r}_n(t,z) - r(t,z)) \\ =&\Big(\sum_{i=1}^{n} f_{ni}(t) - E[f_{ni}(t)]\Big) + \Big(\sum_{i=1}^{n} E[f_{ni}(t)] - \sqrt{nh} r(t,z)\Big) \equiv A_{n1}(t) + A_{n2}(t). \end{aligned}$$

We first claim that $\sup_{t \in \mathcal{T}} |A_{n2}(t)| = o_p(1)$. Note that

$$\begin{aligned} &E[f_{ni}(t)] = E\Big[\frac{1}{\sqrt{nh}} K\big(\frac{Z_i - z}{h}\big) q(W_i, t)\Big] \\ =&\sqrt{\frac{h}{n}} E\Big[\frac{1}{h} K\big(\frac{Z_i - z}{h}\big) q(W_i, t)\Big] \\ =&\sqrt{\frac{h}{n}} \int \frac{1}{h} K\big(\frac{Z_i - z}{h}\big) \mu(t,z) f_z(z) dz \\ =&\sqrt{\frac{h}{n}} \int K(u) r(t, z + uh) du \\ =&\sqrt{\frac{h}{n}} \int K(u) r(t,z) + uhr'(t,z) + u^2 h^2 r''(t, z^*) du \\ =&\sqrt{\frac{h}{n}} (r(t,z) + h^2 \int u^2 K(u) r''(t, z^*) du), \qquad \text{(A.1)} \end{aligned}$$

where z^* denote a point between z and $z + hu$ so that the mean expansion hold. The derivation of (A.1) follows from standard arguments such as (3.49) of Pagan and Ullah (1999, page 99). Therefore,

$$\begin{aligned}\sup_{t\in\mathcal{T}}|A_{n2}(t)| &= \sup_{t\in\mathcal{T}}|A_2(t)| \\ &= \sup_{t\in\mathcal{T}}\Big|\sum_{i=1}^{n}E[f_{ni}(t)] - \sqrt{nh}r(t,z)\Big| \\ &= \sqrt{nh^5}\sup_{t\in\mathcal{T}}\|\int u^2K(u)r''(t,z^*)du\| \\ &\le \sqrt{nh^5}\|\int u^2K(u)M_1du\| \to 0,\end{aligned}$$

for some $M_1 < \infty$ that does not depend on n. The first three equalities hold by definitions, the first inequality holds by Assumption 4.2 and the fact that $r''(t,z) = \mu''(t,z)f(z)+\mu(t,z)f''(z)+2\mu'(t,z)f'(z)$, and the $o(1)$ holds because $nh^5 \to 0$ from Assumption 4.3.

Next, we show that $A_{n1}(t)$ weakly converges to a Gaussian process by checking the conditions of Pollard's FCLT. For condition (i), by Assumption 4.1 (iii) and by Lemma E1 of Andrews and Shi (2013), we have $\{f_{ni}(t):\ t\in\mathcal{T},\ 1\le i\le n,\ n\ge 1\}$ is manageable w.r.t. the envelope functions $\{F_{ni}) : i \le n, 1 \le n\}$ in the sense of Definition 7.9 of Pollard (1990). For condition (ii), first note that

$$\begin{aligned}&\sum_{i=1}^{n}E[f_{ni}(t_1)f_{ni}(t_2)] = nE[f_{ni}(t_1)f_{ni}(t_2)] \\ =&n\Big(E\Big[\frac{1}{nh}K^2\Big(\frac{Z-z}{h}\Big)q(W_n,t_1)q(W_n,t_2)\Big]\Big) \\ =&E\Big[\frac{1}{h}K^2\Big(\frac{Z-z}{h}\Big)q(W_n,t_1)q(W_n,t_2)\Big] \\ =&\int K^2(u)\zeta_n(t_1,t_2,z+hu)f_{n,z}(z+hu)du \\ =&\|K\|_2^2\zeta_n(t_1,t_2,z)f_{n,z}(z) + h\int uK^2(u)\cdot\big(\zeta_n'(t_1,t_2,z^*)f_{n,z}(z^*)+\zeta_n(t_1,t_2,z^*)f_{n,z}'(z^*)\big)du \\ =&\|K\|_2^2\zeta_n(t_1,t_2,z)f_{n,z}(z) + O(h),\end{aligned} \tag{A.2}$$

where $O(h)$ result holds uniformly over $t_1, t_2 \in \mathcal{T}$ and for all n. Also, recall that

$$E[f_{ni}(t_1)] = \sqrt{\frac{h}{n}}\Big(r(t,z) + h^2\int u^2K(u)r''(t,z^*)du\Big),$$

so $\sum_{i=1}^{n}E[f_{ni}(t_1)]E[f_{ni}(t_2)] = n\cdot(E[f_{ni}(t_1)]E[f_{ni}(t_2)]) = O(h)$ and this $O(h)$ result also holds uniformly over $t_1, t_2 \in \mathcal{T}$ and for all n. By these two results, we have

$$E[A_{n1}(t_1)A_{n1}(t_2)] = \|K\|_2^2\zeta_n(t_1,t_2,z)f_{n,z}(z) + O(h) - O(h) \to \|K\|_2^2\zeta(t_1,t_2,z)f,$$

so condition (ii) holds. For condition (iii), note that

$$\begin{aligned}&\sum_{i=1}^{n}E[F_{ni}^2] = nE[F_{ni}^2] \\ =&E\Big[\frac{1}{h}K^2\Big(\frac{Z-z}{h}\Big)Q^2(W_n)\Big] \\ =&\int K^2(u)E[Q^2(W_n)|Z=z+hu]f_{n,z}(z+hu)du \\ \le&\|K\|_2^2M_1du < \infty,\end{aligned}$$

where the first 3 equalities hold by definitions, the first inequality holds because $E[Q^2(W_n)|Z=z]f_{n,z}(z)$ is assumed to be bounded uniformly over n. Therefore, condition (iii) holds. For condition (iv), note that for all $\epsilon > 0$ and for some $\delta > 0$

$$\begin{aligned}
&\sum_{i=1}^{n} E[F_{ni}^2 \cdot 1(F_{ni} > \epsilon)] = nE[F_{ni}^2 \cdot 1(F_{ni} > \epsilon)] \\
\le& nE[F_{ni}^{2+\delta}/\epsilon^{\delta}] \\
=& n\epsilon^{-\delta} E\Big[\sqrt{\frac{1}{(nh)^{2+\delta}}} K^{2+\delta}\Big(\frac{Z-z}{h}\Big) Q^{2+\delta}(W_n)\Big] \\
=& (nh)^{-\delta/2}\epsilon^{-\delta} E\Big[\frac{1}{h} K^{2+\delta}\Big(\frac{Z-z}{h}\Big) Q^{2+\delta}(W_n)\Big] \\
=& (nh)^{-\delta/2}\epsilon^{-\delta} \int K^{2+\delta}(u) E[Q^{2+\delta}(W_n)|Z=z+uh] f_{n,z}(z+uh) du \\
=& (nh)^{-\delta/2}\epsilon^{-\delta} M_1 \to 0,
\end{aligned}$$

for some $M_1 > 0$. The argument is similar to the proof for Theorem 2.1, so condition (iv) holds. For condition (v), note that uniformly over $t_1, t_2 \in \mathcal{T}$,

$$\begin{aligned}
&\rho_n^2(t_1,t_2) = \sum_{i=1}^{n} E[(f_{ni}(t_1) - f_{ni}(t_2))^2] \\
=& nE[(f_{ni}^2(t_1) - 2f_{ni}(t_1)f_{ni}(t_2) + f_{ni}^2(t_2)] \\
=& \|K\|_2^2 \cdot f_{n,z}(z) \cdot (\zeta_n(t_1,t_1,z) + \zeta_n(t_2,t_2,z) + 2\zeta_n(t_1,t_2,z)) + O(h) \\
\to& \|K\|_2^2 \cdot f \cdot (\zeta(t_1,t_1,z) + \zeta(t_2,t_2,z) + 2\zeta(t_1,t_2,z)),
\end{aligned}$$

where the argument is the same as that for condition (ii). This is sufficient to show condition (v). Then by FCLT of Pollard (1990), we have $A_{n1}(\cdot)$ weakly converge to a Gaussian process with covariance kernel equal to $\|K\|_2^2 \cdot f \cdot \zeta(t_1,t_2,z)$. Then this shows that $\sqrt{nh}(\hat{r}_n(t_1,z) - r_n(t_1,z))$ weakly converge to a Gaussian process with covariance kernel equal to $\|K\|_2^2 \zeta(t_1,t_2,z)$.

Then by similar argument, we have $\hat{f}_{n,z}(z) \xrightarrow{p} f$ and $\sqrt{nh}(\hat{f}_{n,z}(z) - f_{n,z}(z)) \xrightarrow{d} N(0, \|K\|_2^2 \cdot f)$.

Next, we have

$$\begin{aligned}
&\sqrt{nh}(\hat{\mu}_n(t,z) - \mu_n(t,z)) = \sqrt{nh}\Big(\frac{\hat{r}_n(t,z)}{\hat{f}_{n,z}(z)} - \frac{r_n(t,z)}{f_{n,z}(z)}\Big) \\
=& \sqrt{nh}\Big(\frac{\hat{r}_n(t,z) - r_n(t,z)}{\hat{f}_{n,z}(z)} + r_n(t,z)\Big(\frac{1}{\hat{f}_{n,z}(z)} - \frac{1}{f_{n,z}(z)}\Big)\Big) \\
=& \sqrt{nh}\Big(\frac{\hat{r}_n(t,z) - r_n(t,z)}{f_{n,z}(z)}\Big) + \sqrt{nh}\Big(\hat{r}_n(t,z) - r_n(t,z)\Big)\Big(\frac{1}{\hat{f}_{n,z}(z)} - \frac{1}{f_{n,z}(z)}\Big) \\
&+ \frac{r_n(t,z)}{f_{n,z}^2(z)}\sqrt{nh}(\hat{f}_{n,z}(z) - f_{n,z}(z)) + \frac{r_n(t,z)}{f_{n,z}(z)}\sqrt{nh}(\hat{f}_{n,z}(z) - f_{n,z}(z))\Big(\frac{1}{\hat{f}_{n,z}(z)} - \frac{1}{f_{n,z}(z)}\Big) \\
=& \sqrt{nh}\Big(\frac{\hat{r}_n(t,z) - r_n(t,z)}{f_{n,z}(z)} - \frac{r_n(t,z)}{f_{n,z}^2(z)}(\hat{f}_{n,z}(z) - f_{n,z}(z))\Big) + o_p(1) \\
=& \sqrt{nh}\frac{1}{f_{n,z}(z)}\Big(\hat{r}_n(t,z) - \mu_n(t,z)\hat{f}_{n,z}(z)\Big) + o_p(1) \\
=& \sqrt{\frac{1}{nh}}\frac{1}{f_{n,z}(z)}\sum_{i=1}^{n} K\Big(\frac{Z_{ni}-z}{h}\Big)\Big(q(W_{ni},t) - \mu_n(t,z)\Big) + o_p(1) \qquad \text{(A.3)}
\end{aligned}$$

where the first three equalities hold by rewriting, the fourth equality holds because the second term and the fourth terms are both $o_p(1)$ uniformly over $t \in \mathcal{T}$, the fifth equality holds by rewriting and the last equality holds once we expand $\hat{r}_n(t,z)$ and $\hat{f}_{n,z}(z)$. Then by the same argument for $\sqrt{nh}(\hat{r}_n(t,z)-r(t,z))$, we can show that $\sqrt{nh}(\hat{\mu}_n(t,z)-\mu_n(t,z)) \Rightarrow \Phi_{h_2}$ where $h_2 = \lim_{n\to\infty} h_{2,P_n,z}$.

To show the second part, note that the proof is very similar to that for Theorem 2.1. Define

$$\begin{aligned}
\hat{\rho}_n(t_1,t_2,z) &= \frac{1}{nh}\frac{1}{\hat{f}^2_{n,z}(z)}\sum_{i=1}^n K^2\Big(\frac{Z_i-z}{h}\Big)(q(W_i,t_1)-\hat{\mu}_n(t_1,z))(q(W_i,t_2)-\hat{\mu}_n(t_2,z)),\\
\ddot{\rho}_n(t_1,t_2,z) &= \frac{1}{nh}\frac{1}{f^2_{n,z}(z)}\sum_{i=1}^n K^2\Big(\frac{Z_i-z}{h}\Big)(q(W_i,t_1)-\mu_n(t_1,z))(q(W_i,t_2)-\mu_n(t_2,z)),\\
E_{P_n}[\ddot{\rho}_n(t_1,t_2,z)] &= E_{P_n}\Big[\frac{1}{h}\frac{1}{f^2_{n,z}(z)}K^2\Big(\frac{{}_iZ-z}{h}\Big)(q(W_i,t_1)-\mu_n(t_1,z))(q(W_i,t_2)-\mu_n(t_2,z))\Big],\\
\hat{\Lambda}_n(z) &= \frac{1}{nh}\sum_{i=1}^n K\Big(\frac{Z_i-z}{h}\Big)Q(W_i)\Big/\frac{1}{nh}\sum_{i=1}^n K\Big(\frac{Z_i-z}{h}\Big),\\
\Lambda_n(z) &= E[Q(W_i)|Z=z].
\end{aligned}$$

We first use Andrews and Shi's (2015, Lemma AN4(b)) result to show that $\sup_{t_1,t_2\in\mathcal{T}}\big|\hat{\rho}_n(t_1,t_2,z)-h_2(t_1,t_2)\big| \xrightarrow{p} 0$. Note that

$$\begin{aligned}
&\hat{\rho}_n(t_1,t_2,z)-h_2(t_1,t_2)\\
=&\big(\hat{\rho}_n(t_1,t_2,z)-\ddot{\rho}_n(t_1,t_2,z)\big)+\big(\ddot{\rho}_n(t_1,t_2,z)-E_{P_n}[\ddot{\rho}_n(t_1,t_2,z)]\big)\\
&+\big(E_{P_n}[\ddot{\rho}_n(t_1,t_2,z)]-h_{2,P_n}(t_1,t_2,z)\big)+\big(h_{2,P_n}(t_1,t_2,z)-h_2(t_1,t_2)\big)\\
\equiv& B_{1n}(t_1,t_2)+B_{2n}(t_1,t_2)+B_{3n}(t_1,t_2)+B_{4n}(t_1,t_2).
\end{aligned}$$

Note that $\sup_{t_1,t_2\in\mathcal{T}}|B_{4n}(t_1,t_2)| \to 0$ by Assumption 4.1(vii). $\sup_{t_1,t_2\in\mathcal{T}}|B_{3n}(t_1,t_2)| \to 0$ holds by the same argument for (A.2). $\sup_{t_1,t_2\in\mathcal{T}}|B_{1n}(t_1,t_2)| \xrightarrow{p} 0$ because $\hat{f}_{n,z}(z) \xrightarrow{p} f$ and $\sup_{t\in\mathcal{T}}|\hat{\mu}_n(t,z)-\mu_n(t,z)| \xrightarrow{p}$ from first part. Under Assumption 4.1, $\sup_{t_1,t_2\in\mathcal{T}}|B_{2n}(t_1,t_2)| \xrightarrow{p} 0$ follows form the same arguments in (12.24)-(12.26) of Andrews and Shi (2015b). These are sufficient to show that $\sup_{t_1,t_2\in\mathcal{T}}\big|\hat{\rho}_n(t_1,t_2,z)-h_2(t_1,t_2)\big| \xrightarrow{p} 0$.

Under Assumption 4.1, we have

$$\begin{aligned}
&\Big|\frac{1}{nh}\sum_{i=1}^n K^2\Big(\frac{Z_i-z}{h}\Big)(Q(W_i)+\hat{\Lambda}_n(z))^2-\|K\|_2^2\cdot f_{n,z}(z)\cdot E_{P_n}\Big[(Q(W_i)+\Lambda_n(z))^2|Z_i=z\Big]\Big|\\
\le&\Big|\frac{1}{nh}\sum_{i=1}^n K^2\Big(\frac{Z_i-z}{h}\Big)(Q(W_i)+\hat{\Lambda}_n(z))^2-\frac{1}{nh}\sum_{i=1}^n K^2\Big(\frac{Z_i-z}{h}\Big)(Q(W_i)+\Lambda_n(z))^2\Big|\\
&+\Big|\frac{1}{nh}\sum_{i=1}^n K^2\Big(\frac{Z_i-z}{h}\Big)(Q(W_i)+\Lambda_n(z))^2-\|K\|_2^2\cdot f_{n,z}(z)\cdot E_{P_n}\Big[(Q(W_i)+\Lambda_n(z))^2|Z_i=z\Big]\Big|\\
&\equiv C_{n1}+C_{n2}.
\end{aligned}$$

Note that $C_{n1} \xrightarrow{p} 0$ because that $|\hat{\Lambda}_n(z)-\Lambda_n(z)| \xrightarrow{p} 0$, and $C_{n2} \xrightarrow{p} 0$ by the same argument for A.2.

Similarly, we have

$$\Big|\frac{1}{nh}\sum_{i=1}^{n}K^{2+\delta}\Big(\frac{Z_i-z}{h}\Big)(Q(W_i)+\hat{\Lambda}_n(z))^{2+\delta}$$

$$-\|K\|_{2+\delta}^{2+\delta}\cdot f_{n,z}(z)\cdot E_{P_n}\Big[(Q(W_i)+\Lambda_n(z))^{2+\delta}|Z_i=z\Big]\Big|\overset{p}{\to}0,$$

where $\|K\|_{2+\delta}^{2+\delta}=\int K(u)^{2+\delta}du$. By assumptions, we have

$$\limsup\|K\|_2^2\cdot f_{n,z}(z)\cdot E_{P_n}\Big[(Q(W_i)+\Lambda_n(z))^2|Z_i=z\Big]\le M,$$

$$\limsup\|K\|_{2+\delta}^{2+\delta}\cdot f_{n,z}(z)\cdot E_{P_n}\Big[(Q(W_i)+\Lambda_n(z))^{2+\delta}|Z_i=z\Big]\le M,$$

for some $M<\infty$.

Therefore, for any subsequence $\{b_n\}$, there exists a further subsequence $\{k_n\}$ such that

$$\sup_{t_1,t_2\in\mathcal{T}}\big|\hat{\rho}_{k_n}(t_1,t_2,z)-h_2(t_1,t_2)\big|\overset{a.s.}{\to}0.$$

$$\limsup\Big|\frac{1}{nh}\sum_{i=1}^{n}K^2\Big(\frac{Z_i-z}{h}\Big)(Q(W_i)+\hat{\Lambda}_n(z))^2\Big|\le M \text{ almost surely, and}$$

$$\limsup\Big|\frac{1}{nh}\sum_{i=1}^{n}K^{2+\delta}\Big(\frac{Z_i-z}{h}\Big)(Q(W_i)+\hat{\Lambda}_n(z))^{2+\delta}\Big|\le M \text{ almost surely.}$$

Therefore, by similar argument for the second part of Theorem 2.1, we have $\widehat{\Phi}_{k_n}^u(\cdot)\overset{a.s.}{\Rightarrow}\Phi_{h_2}(\cdot)$ and this shows $\widehat{\Phi}_n^u(\cdot)\overset{p}{\Rightarrow}\Phi_{h_2}(\cdot)$. □

References

Andrews, D. W. K. (1994). "Empirical process methods in econometrics," Handbook of Econometrics, in: R. F. Engle and D. McFadden (ed.), Vol. 4, 2247–2294, Elsevier.

Andrews, D. W. K., and X. Shi (2013). "Inference Based on Conditional Moment Inequalities," *Econometrica*, **81**, 609–666.

Andrews, D. W. K., and X. Shi (2014). "Nonparametric Inference Based on Conditional Moment Inequalities," *Journal of Econometrics*, **179**, 31–45.

Barrett, G. F., and S. G. Donald (2003). "Consistent Tests for Stochastic Dominance," *Econometrica*, **71**, 71–104.

Bugni, F. A., I. A. Canay, and X. Shi (2015). "Specification Tests for Moment Inequality models," *Journal of Econometrics*, **185**, 259–282.

Donald, S. G., and Y.-C. Hsu (2014). "Estimation and Inference for Distribution Functions and Quantile Functions in Treatment Effect Models," *Journal of Econometrics*, **178**, 383–397.

Donald, S. G., and Y.-C. Hsu (2013). "Improving the Power of Tests of Stochastic Dominance," *Econometric Reviews*, forthcoming.

Hsu, Y.-C. (2015). "Consistent Tests for Conditional Treatment Effects," Working Paper.

Hsu, Y.-C., K. Kan, and T.-C. Lai (2015). "Distribution and Quantile Structural Functions in Treatment Effect Models: Application to Smoking Effects on Wages," Working Paper.

Hsu, Y.-C., T.-C. Lai, and R. P. Lieli (2015). "Estimation and Inference for Distribution Functions and Quantile Functions in Endogenous Treatment Effect Models," Working Paper.

Hsu, Y.-C., C.-A. Liu, and X. Shi (2015). "Testing for Generalized Regression Monotonicity," Working Paper.

Kaido, H. and A. Santos (2014). "Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities," *Econometrica*, **82**, 387–413.

Kline. P. and A. Santos (2012). "A Score Based Approach to Wild Bootstrap Inference," *Journal of Econometric Methods*, **1**, 23–41.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*, Springer Science+Business Media.

Pollard, D. (1990). *Empirical Processes: Theory and Application, CBMS Conference Series in*

Probability and Statistics, Vol. 2. Hayward, CA: Institute of Mathematical Statistics.

van der Vaart, A. W., and J. A. Wellner (1996). *Weak Convergence and Empirical Processes: With Application to Statistics*, New York: Springer-Verlag.