On Rosenbaum's Rank-based Matching Estimator

Matias D. Cattaneo[*]   Fang Han[†]   Zhexiao Lin[‡]

December 12, 2023

Abstract

In two influential contributions, Rosenbaum (2005, 2020) advocated for using the distances between component-wise ranks, instead of the original data values, to measure covariate similarity when constructing matching estimators of average treatment effects. While the intuitive benefits of using covariate ranks for matching estimation are apparent, there is no theoretical understanding of such procedures in the literature. We fill this gap by demonstrating that Rosenbaum's rank-based matching estimator, when coupled with a regression adjustment, enjoys the properties of double robustness and semiparametric efficiency without the need to enforce restrictive covariate moment assumptions. Our theoretical findings further emphasize the statistical virtues of employing ranks for estimation and inference, more broadly aligning with the insights put forth by Peter Bickel in his 2004 Rietz lecture (Bickel, 2004).

Keywords: rank-based statistics, matching estimators, average treatment effect, regression adjustment, semiparametric efficiency.

1 Introduction

Consider the problem of estimating the average treatment effect (ATE),

$$\tau := \mathrm{E}\Big[Y(1) - Y(0)\Big],$$

based on an observational study encompassing n observations of a binary treatment $D \in \{0, 1\}$, some measured pre-treatment covariates $X \in \mathbb{R}^d$, and an outcome $Y = Y(D) \in \mathbb{R}$ that is realized from the two potential outcomes $(Y(0), Y(1))$. Among the techniques employed to estimate τ, nearest neighbor (NN) matching stands as one of the most widely adopted and comprehensible approaches (see Stuart, 2010, and references therein). These estimators aim to impute the missing potential outcome of each unit in one treatment group by finding units from the opposite treatment group whose covariate profile closely resemble that of the unit with the missing potential outcome. The quantification of covariate similarity relies on the user-specified *distance metric* between the (distribution of the) covariates of each unit.

---

[*]Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA; e-mail: `cattaneo@princeton.edu`

[†]Department of Statistics, University of Washington, Seattle, WA 98195, USA; e-mail: `fanghan@uw.edu`

[‡]Department of Statistics, University of California, Berkeley, CA 94720, USA; e-mail: `zhexiaolin@berkeley.edu`

Abadie and Imbens (2006) laid out the mathematical groundwork to study NN matching estimators employing the Euclidean distance metric, while Abadie and Imbens (2011) established a root-n central limit theorem for a bias-corrected version of those NN matching estimators. More recently, Lin et al. (2023) established connections between NN estimators and augmented inverse probability weighted (AIPW) methods (Robins et al., 1994; Scharfstein et al., 1999), thereby establishing double robustness and semiparametric efficiency theory for NN matching estimators when the number of matches diverges to infinity with the sample size.

The aforementioned theoretical work, however, centers around the Euclidean distance metric for determining NN matches. This approach may exhibit sensitivity to alterations in scale and to the presence of extreme outliers or heavy-tailed distributions. Indeed, all the existing theoretical results on matching require a compact supported covariates, which is theoretically hard to alleviate, if not impossible. On the other hand, in practice, distance metrics are often derived from a "standardized" representation of the data, and the selection of a distance metric is an important factor in causal inference because various metrics can lead to different conclusions (Rosenbaum, 2020, Chapter 9).

This paper focuses on a particular standardization approach that identifies NNs by measuring the Euclidean distance between the component-wise ranks of the covariates X, as proposed for the celebrated Rosenbaum's rank-based matching estimator (Rosenbaum, 2020, Chapter 9.3). The concept of rank-based standardization is straightforward to interpret, easy to implement, and computationally efficient, while also being scale-invariant and insensitive to heavy-tailed distributions. Furthermore, due to their data adaptivity, rank-based methods are often used in other treatment effect settings such as for analysis of experiments (Rosenbaum, 2020), regression discontinuity plots (Calonico et al., 2015), and binscatter regressions (Cattaneo et al., 2023).

The challenge in formally studying rank-based methods lies on the theoretical side, as the transformation of the covariates into their ranks disrupts the independence structure of the original data, and thus complicates the subsequent statistical analysis. Our main theorem (Theorem 3.1) offers the first theoretical analysis of Rosenbaum's rank-based matching estimator, elucidating many of its appealing properties when combined with regression adjustments. In particular, our theory not only confirms Rosenbaum's intuition that the rank-based distance can limit the influence of outliers and heavy-tailed distributions (Rosenbaum, 2020, page 210), but also demonstrates that the rank-based matching estimator can be doubly robust and semiparametrically efficient, particularly without imposing restrictive moment assumptions on the distribution of the covariates. More broadly, they align with Peter Bickel's 2004 Rietz lecture advocating for "standardization by ranks" when performing statistical and machine learning related tasks (Bickel, 2004).

Our paper further offers two technical contributions as a by-product of the analysis, which may be of independent interest. First, Theorem 3.1 establishes consistency, asymptotic linearity, and semiparametric efficiency of Rosenbaum's Rank-based Matching Estimator under generic high-level conditions on the regression adjustment. The proof of that theorem relies on a careful combination of empirical process theory for rank-based statistics and tools established in Lin and Han (2022) and Lin et al. (2023) for matching estimators involving a growing number of nearest neighbors, which are generalized herein to accommodate standardization/transformation functions, of which

component-wise ranking is one particular example. Second, Theorem 4.1 presents novel mean square and uniform convergence rates for series estimators when the covariates are generated via possibly unknown functions of the original independent variables, of which component-wise ranking is one particular example. Those results are proven for general series estimators (Newey, 1997; Belloni et al., 2015) with covariate-generated conditioning variables, and thus lead to suboptimal uniform approximation results, but we also discuss how they may be upgraded to deliver optimal uniform convergence rates for partition-based series estimators (Huang, 2003; Cattaneo and Farrell, 2013; Cattaneo et al., 2020, 2023).

Paper organization. The rest of this paper is organized as follows. In Section 2 we describe the problem setup and introduce the studied rank-based matching estimator. Section 3 establishes the main result of this paper under general regression adjustments. Section 4 discusses primitive conditions for regression adjustment using series estimation. Section 5 generalizes the results in Section 3. All proofs are in the appendix.

2 Setup

We adopt the standard potential outcomes causal model for a binary treatment, where it is assumed that there are n independent and identically (i.i.d.) distributed realizations $\{X_i, D_i, Y_i(0), Y_i(1)\}_{i=1}^n$, of a quadruple $(X, D, Y(0), Y(1))$. In practice, we are only able to observe a part of the data, i.e., $\{X_i, D_i, Y_i := Y_i(D_i)\}_{i=1}^n$. The goal is to conduct estimation and inference for the population ATE,

$$\tau = \mathrm{E}[Y(1) - Y(0)],$$

based only on the observed data.

Rosenbaum's rank-based matching approach estimates τ by plugging the component-wise ranks of X_i's, instead of the original values, into the NN matching mechanism, with each unit matched to M units in the opposite treatment group with replacement. It proceeds in three step as follows.

Step 1. Given a sample $\{(X_i, D_i, Y_i)\}_{i=1}^n$ with $X_i = (X_{i,1}, \ldots, X_{i,d})^\top$, introduce the vector of marginal empirical cumulative distribution functions (CDFs), $\widehat{\boldsymbol{F}}_n(\cdot) : \mathbb{R}^d \to [0,1]^d$, such that for any input $x = (x_1, \ldots, x_d)^\top \in \mathbb{R}^d$,

$$\widehat{\boldsymbol{F}}_n(x) := (\widehat{F}_{n,1}(x_1), \ldots, \widehat{F}_{n,d}(x_d))^\top, \;\; \text{with} \;\; \widehat{F}_{n,k}(x_k) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_{i,k} \leq x_k), \;\; k \in [\![d]\!].$$

Here $\mathbb{1}(\cdot)$ stands for the indicator function and $[\![d]\!] := \{1, 2, \ldots, d\}$. For each $i \in [\![n]\!]$, introduce $\widehat{U}_i := \widehat{\boldsymbol{F}}_n(X_i) \in \mathbb{R}^d$. It is understood that each component of $n\widehat{U}_i$ represents the corresponding rank of $X_{i,\cdot}$ among $X_{1,\cdot}, \ldots, X_{n,\cdot}$. The vector of marginal population CDFs is $\boldsymbol{F}(\cdot) : \mathbb{R}^d \to [0,1]^d$, such that for any input $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$,

$$\boldsymbol{F}(x) := (F_1(x_1), \ldots, F_d(x_d))^\top, \;\; \text{with} \;\; F_k(x_k) := \mathrm{P}(X_{1,k} \leq x_k), \;\; k \in [\![d]\!].$$

Define $U := \boldsymbol{F}(X) \in [0,1]^d$ and for each $i \in [\![n]\!]$, $U_i := \boldsymbol{F}(X_i) \in [0,1]^d$.

Step 2. To correct for the estimation bias from matching, regression adjustment is employed. Let

$\widehat{\mu}_0(\cdot)$ and $\widehat{\mu}_1(\cdot)$ be mappings from $[0,1]^d$ to $\mathbb{R}$ such that they separately estimate the conditional means of the outcomes,

$$\mu_0(u) := \mathrm{E}[Y \mid U = u, D = 0] \qquad \text{and} \qquad \mu_1(u) := \mathrm{E}[Y \mid U = u, D = 1].$$

We obtain $\widehat{\mu}_0$ and $\widehat{\mu}_1$ by regressing Y_i's in either the treatment or control group on the corresponding $\{\widehat{U}_i\}$'s, respectively.

Step 3. Implement bias-corrected nearest neighbor matching on $(\widehat{U}_i, D_i, Y_i)$'s. Specifically, let $\mathcal{J}(i)$ represent the index set of the M-NNs of $\widehat{U}_i$ in $\{\widehat{U}_j : D_j = 1 - D_i\}_{j=1}^n$, measured using the Euclidean metric $\|\cdot\|$ with *ties broken in arbitrary way*. The final rank-based matching estimator is

$$\widehat{\tau} := \frac{1}{n} \sum_{i=1}^{n} \left[\widehat{Y}_i(1) - \widehat{Y}_i(0) \right], \tag{2.1}$$

where, for $\omega \in \{0,1\}$,

$$\widehat{Y}_i(\omega) := \begin{cases} Y_i, & \text{if } D_i = \omega, \\ \frac{1}{M} \sum_{j \in \mathcal{J}(i)} (Y_j + \widehat{\mu}_\omega(\widehat{U}_i) - \widehat{\mu}_\omega(\widehat{U}_j)), & \text{if } D_i = 1 - \omega \end{cases}.$$

3 Main Result

Following the notation system of Abadie and Imbens (2006, 2011), let $K(i)$ represent the number of matched times for each unit i, i.e.,

$$K(i) := \sum_{j=1, D_j = 1 - D_i}^{n} \mathbb{1}\big(i \in \mathcal{J}(j) \big).$$

In this paper, however, $K(i)$ denotes the matched times according to *the rank-based distance*, not the original Euclidean distance. Following Lin and Han (2022) and Lin et al. (2023), the rank-based bias-corrected matching estimator in (2.1) can be represented as an AIPW estimator:

$$\widehat{\tau} = \widehat{\tau}^{\text{reg}} + \frac{1}{n} \sum_{i=1}^{n} (2D_i - 1)\Big(1 + \frac{K(i)}{M}\Big)\widehat{R}_i, \tag{3.1}$$

where

$$\widehat{\tau}^{\text{reg}} := \frac{1}{n} \sum_{i=1}^{n} [\widehat{\mu}_1(\widehat{U}_i) - \widehat{\mu}_0(\widehat{U}_i)] \quad \text{and} \quad \widehat{R}_i := Y_i - \widehat{\mu}_{D_i}(\widehat{U}_i), \text{ for } i \in \llbracket n \rrbracket.$$

We employ this insight throughout our large sample distributional analysis of $\widehat{\tau}$.

To establish our main theorem we impose some assumptions. The first assumption is posed to regulate basic features of the data generating distribution, in particular making the estimation problem identifiable. Conceptually, the main difference with prior literature is that the assumption concerns the triple (U_i, D_i, Y_i)'s, that is, X_i is replaced by U_i, the scaled population rank.

Assumption 3.1.

(i) *For almost all $u \in [0,1]^d$, D is independent of $(Y(0), Y(1))$ conditional on $U = u$, and there exists a fixed constant $c > 0$ such that $c < e(u) := \mathrm{P}(D = 1 \mid U = u) < 1 - c$.*

(ii) $[(X_i, D_i, Y_i)]_{i=1}^n$ *are i.i.d. following the joint distribution of* (X, D, Y).

(iii) $\mathrm{E}[(Y(\omega) - \mu_\omega(U))^2 \,|\, U = u]$ *is uniformly bounded for almost all* $u \in [0,1]^d$ *and* $\omega = 0, 1$.

(iv) $\mathrm{E}[\mu_\omega^2(U)]$ *is bounded for* $\omega = 0, 1$.

The second assumption ensures that the regression adjustment procedure is, at least, well-posited in the sense that the estimator $\widehat{\mu}_\omega(x)$ is uniformly consistent for some well-behaved (possibly misspecified) conditional expectation. Let $\|\cdot\|_\infty$ be the L^∞ function norm.

Assumption 3.2. *For* $\omega = 0, 1$, *there exists a deterministic, possibly changing with n, continuous function* $\bar{\mu}_\omega(\cdot) : [0,1]^d \to \mathbb{R}$ *such that* $\mathrm{E}[\bar{\mu}_\omega^2(U)]$ *is uniformly bounded and the estimator* $\widehat{\mu}_\omega(x)$ *satisfies* $\|\widehat{\mu}_\omega - \bar{\mu}_\omega\|_\infty = o_{\mathrm{P}}(1)$.

The next assumption regulates the population rank-transformed random vector U, which identifies the copula distribution for X (Joe, 2014). This assumption requires X to be continuous, but discrete components of X can be easily handled by conditioning (Stuart, 2010).

Assumption 3.3. *The Lebesgue density of U exists and is continuous over its support.*

The next three assumptions concern the case of consistent population regression functions $\mu_0(U)$ and $\mu_1(U)$, and will be used in contrast to Assumption 3.2, where the regression adjustment procedure is allowed to be inconsistent for the population ranked-based regression functions. More precisely, Assumption 3.2 vis-á-vis Assumptions 3.4–3.6 are used for establishing the double robustness and semiparametric efficiency of $\widehat{\tau}$, respectively. Using standard multi-index notation, let Λ_k is the set of all d-dimensional vectors of nonnegative integers $t = (t_1, \ldots, t_d)$ such that $|t| = \sum_{i=1}^d t_i = k$ with k any positive integer, and $\partial^t \mu_\omega$ denotes the corresponding partial derivative of μ_ω.

Assumption 3.4. *For* $\omega = 0, 1$, μ_ω *is continuous and the estimator* $\widehat{\mu}_\omega(x)$ *satisfies* $\|\widehat{\mu}_\omega - \mu_\omega\|_\infty = o_{\mathrm{P}}(1)$.

Assumption 3.5.

(i) $\mathrm{E}[(Y(\omega) - \mu_\omega(U))^2 \,|\, U = u]$ *is uniformly bounded away from zero for almost all* $u \in [0,1]^d$ *and* $\omega = 0, 1$.

(ii) *There exists a constant* $c > 0$ *such that* $\mathrm{E}[|Y(\omega) - \mu_\omega(U)|^{2+c} \,|\, U = u]$ *is uniformly bounded for almost all* $u \in [0,1]^d$ *and* $\omega = 0, 1$.

(iii) $\max_{t \in \Lambda_{\max\{\lfloor d/2 \rfloor, 1\}+1}} \|\partial^t \mu_\omega\|_\infty$ *is bounded, where* $\lfloor \cdot \rfloor$ *stands for the floor function.*

Assumption 3.6. *For* $\omega = 0, 1$, *the estimator* $\widehat{\mu}_\omega(x)$ *satisfies*

$$\max_{t \in \Lambda_{\max\{\lfloor d/2 \rfloor, 1\}+1}} \|\partial^t \widehat{\mu}_\omega\|_\infty = O_{\mathrm{P}}(1) \quad \text{and} \quad \max_{t \in \Lambda_\ell} \|\partial^t \widehat{\mu}_\omega - \partial^t \mu_\omega\|_\infty = O_{\mathrm{P}}(n^{-\gamma_\ell}) \text{ for all } \ell \in [\![\max\{\lfloor d/2 \rfloor, 1\}]\!],$$

with some constants γ_ℓ's *satisfying* $\gamma_\ell > \max\{\frac{1}{2} - \frac{\ell}{d}, 0\}$ *for* $\ell = 1, 2, \ldots, \max\{\lfloor d/2 \rfloor, 1\}$.

Several remarks on the above assumptions align with our conceptual discussion. First, as the quantile transformation preserves all information, Assumption 3.1 is either equivalent to, or weaker

than, the standard assumptions in the matching estimation literature. Second, Assumption 3.2 accommodates regression model misspecification, and its validity may be verified by leveraging the fundamental projection principles underlying regression techniques (see, also, the discussions in Section 4). Third, Assumption 3.3 constitutes a mild condition, notably satisfied by distribution families such as the Gaussian (copula) and Cauchy (copula). Lastly, Assumptions 3.4 through 3.6 merit more discussion: due to the shift from using X_i's as inputs to $\widehat{U}_i$'s in the regression function, direct verification using standard results from the nonparametric smoothing estimation literature is no longer possible. We return to this technical issue in Section 4, where we consider explicitly least squares series estimation (Newey, 1997; Huang, 2003; Cattaneo and Farrell, 2013; Belloni et al., 2015; Cattaneo et al., 2020) to illustrate the verification of Assumption 3.2 and Assumptions 3.4–3.6.

We are now ready to present our main theorem for Rosenbaum's rank-based matching estimator.

Theorem 3.1 (Main Theorem).

(i) *(Double robustness of $\widehat{\tau}$) If either Assumptions 3.1, 3.2, 3.3, $M \log n/n \to 0$, and $M \to \infty$ as $n \to \infty$ hold, or Assumptions 3.1 and 3.4 hold, then*

$$\widehat{\tau} - \tau \xrightarrow{\mathsf{p}} 0.$$

(ii) *(Semiparametric efficiency of $\widehat{\tau}$) Assume Assumptions 3.1, 3.3, 3.5, 3.6 hold. Define*

$$\gamma = \max\left\{\left[1 - \frac{1}{2}\frac{d}{\max\{\lfloor d/2 \rfloor, 1\} + 1}\right], \min_{\ell \in \llbracket \max\{\lfloor d/2 \rfloor, 1\}\rrbracket}\left[1 - \left(\frac{1}{2} - \gamma_\ell\right)\frac{d}{\ell}\right]\right\},$$

where the $(\gamma_\ell : \ell = 1, 2, \ldots, \max\{\lfloor d/2 \rfloor, 1\})$ are introduced in Assumption 3.6. If $M \to \infty$ and $M/n^\gamma \to 0$ as $n \to \infty$, then

$$\sqrt{n}(\widehat{\tau} - \tau) \xrightarrow{\mathsf{d}} N(0, \sigma^2),$$

where

$$\sigma^2 := \mathrm{E}\left[\mu_1(U) - \mu_0(U) + \frac{D(Y - \mu_1(U))}{e(U)} - \frac{(1 - D)(Y - \mu_0(U))}{1 - e(U)} - \tau\right]^2.$$

(iii) *(Variance Estimation) If the conditions in part (ii) and Assumption 3.4 hold, then*

$$\widehat{\sigma}^2 - \sigma^2 \xrightarrow{\mathsf{p}} 0$$

where

$$\widehat{\sigma}^2 := \frac{1}{n}\sum_{i=1}^{n}\left[\widehat{\mu}_1(\widehat{U}_i) - \widehat{\mu}_0(\widehat{U}_i) + (2D_i - 1)\left(1 + \frac{K_M(i)}{M}\right)\widehat{R}_i - \widehat{\tau}\right]^2.$$

This theorem establishes three main results. Part (i) shows that the generic bias-corrected Rosenbaum's rank-based matching estimator is doubly robust for a fairly large class of regression estimators based on estimated ranks of the covariates. The result in part (ii) gives general regularity conditions guaranteeing asymptotic normality of the estimator. It follows directly from the second result that the estimator is semiparametrically efficient for estimating the ATE (Hahn, 1998). Finally, part (iii) establishes consistency of the plug-in variance estimator under nearly the same conditions as required for consistency and asymptotic normality.

4 Regression Adjustment using Series Least Squares

The only remaining issue concerning Theorem 3.1 revolves around the use of pairs $(\widehat{U}_i, Y_i)$ for bias correction, as opposed to (X_i, Y_i), or the idealized oracle (U_i, Y_i) pairs. The dependence among the estimated rank-adjusted $\widehat{U}_1, \ldots, \widehat{U}_n$ poses a challenge, making it hard to apply existing results in nonparametric statistics for the direct verification of Assumption 3.2 or Assumptions 3.4–3.6. This section illustrates how these assumptions can be verified when using series least squares regression estimation, covering both canonical approximating functions (e.g., power series, fourier series, splines, wavelets, and piecewise polynomials) as well as general covariate transformations (e.g., high-dimensional least squares regression with structured regressors).

The main result in this section concerns general estimated transformations of the independent variables, based on the underlying regressors only, and allowing for possible misspecification in both fixed-dimension and increasing-dimension least squares regression settings. Thus, we consider a more general setup where we either observe or have approximate information about n i.i.d. pairs $(Y_1, W_1), \ldots, (Y_n, W_n)$ of (Y, W), and the goal is to estimate the conditional expectation

$$\psi(w) := \mathrm{E}[Y \,|\, W = w],$$

using only the outcome variables $\{Y_i\}_{i=1}^n$ and the *generated covariates* $\{\widehat{W}_i\}_{i=1}^n$ that are "approximately close" to $\{W_i\}_{i=1}^n$, where $\widehat{W}_i$'s only depend on $\{W_i\}_{i=1}^n$ but not the outcome variables. To be more precise, we assume throughout that $\widehat{W}_i = g_i(W_1, W_2, \ldots, W_n)$ for some non-random, possibly unknown, functions g_i, each observation $i = 1, 2, \cdots, n$.

Let $p_K(w) = (p_{1K}(w), \ldots, p_{KK}(w))^\top$ be a K-dimensional vector of basis functions so that their linear combination may approximate $\psi(\cdot)$ well when K is sufficiently large, at least under some specific assumptions. However, a good approximation is not strictly required, as we also consider misspecified regression adjustments. This is important in the context of this paper because Theorem 3.1 established a double-robust property of the regression adjusted Rosenbaum's rank-based matching estimator, thereby allowing for misspecified or inconsistent estimators $(\widehat{\mu}_0, \widehat{\mu}_1)$ of the regression functions (μ_0, μ_1). Furthermore, we also allow for the possibility of W having a Lebesgue density that may not be bounded away from zero, as it occurs when its support is unbounded. In our application, due to the rank transformation, the support of the rank-based covariates is bounded but their density may not be bounded away from zero in, e.g., the Gaussian copula case.

The following assumption summarizes our setup. Let $\lambda_{\min}(\cdot)$ be the smallest eigenvalue of the input matrix.

Assumption 4.1.

(i) $(Y_1, W_1), \ldots, (Y_n, W_n)$ *are i.i.d. draws from* $(Y, W) \in \mathcal{Y} \times \mathcal{W} \subseteq \mathbb{R} \times \mathbb{R}^d$.

(ii) $\mathrm{Var}(Y|W)$ *is uniformly bounded and* $\mathrm{E}[Y^2]$ *is bounded.*

(iii) $\widehat{W}_i = g_i(W_1, W_2, \ldots, W_n)$ *for non-random, possibly unknown, functions* $g_i : \mathcal{W}^n \to \mathcal{W}$, $i = 1, 2, \cdots, n$.

(iv) $\lambda_{\min}\big(\mathrm{E}[p_K(W_1)p_K(W_1)^\top]\big) > 0$ *for all K and n.*

The series estimator with generated covariates is

$$\widehat{\psi}_K(w) = p_K(w)^\top \widehat{\beta}_K, \qquad \widehat{\beta}_K \in \operatorname*{argmin}_{b \in \mathbb{R}^K} \frac{1}{n} \sum_{i=1}^n (Y_i - p_K(\widehat{W}_i)^\top b)^2,$$

which gives $\widehat{\beta}_K := (\boldsymbol{P}_n^\top \boldsymbol{P}_n)^- \boldsymbol{P}_n^\top \boldsymbol{Y}$, where $\boldsymbol{P}_n = (p_K(\widehat{W}_1), \ldots, p_K(\widehat{W}_n))^\top \in \mathbb{R}^{n \times k}$, $\boldsymbol{Y} := (Y_1, \ldots, Y_n)^\top$, and $\boldsymbol{A}^-$ denotes a generalized inverse of the matrix $\boldsymbol{A}$. Let $\boldsymbol{P} := (p_K(W_1), \ldots, p_K(W_n))^\top$ be what $\boldsymbol{P}_n$ shall approximate, and

$$\beta_K := \operatorname*{argmin}_{b \in \mathbb{R}^K} \mathrm{E}[(Y_1 - p_K(W_1)^\top b)^2] = \operatorname*{argmin}_{b \in \mathbb{R}^K} \mathrm{E}[(\psi(W_1) - p_K(W_1)^\top b)^2]$$

be what $\widehat{\beta}_K$ shall approximate. It follows that $\psi_K(w) := p_K(w)^\top \beta_K$ is the best L^2 approximation of $\psi(w)$ based on $p_K(w)$, where $\beta_K = \boldsymbol{Q}^- \mathrm{E}[p_K(W_1)\psi(W_1)]$ with $\boldsymbol{Q} := \mathrm{E}[p_K(W_1)p_K(W_1)^\top] \in \mathbb{R}^{K \times K}$.

Our results rely on the following quantities characterizing different aspects of the series estimator and the approximation errors:

$$\lambda_K := \lambda_{\min}(\boldsymbol{Q}), \qquad\qquad \zeta_{q,K} := \max_{t \in \Lambda_q} \sup_{w \in \mathcal{W}} \|\partial^t p_K(w)\|,$$

$$\xi_K^2 := \mathrm{E}[(\psi(W_1) - \psi_K(W_1))^2], \qquad\qquad \vartheta_{q,K} := \max_{t \in \Lambda_q} \|\partial^t \psi - \partial^t \psi_K\|_\infty,$$

and

$$R_n := \|\boldsymbol{\Psi} - \boldsymbol{\Psi}_n\|^2/n, \qquad B_n := \|(\boldsymbol{P} - \boldsymbol{P}_n)\boldsymbol{Q}^{-1/2}\|_2^2/n,$$

where $\boldsymbol{\Psi} := (\psi(W_1), \ldots, \psi(W_n))^\top \in \mathbb{R}^n$, $\boldsymbol{\Psi}_n := (\psi(\widehat{W}_1), \ldots, \psi(\widehat{W}_n))^\top \in \mathbb{R}^n$, and $\| \cdot \|_2$ denotes the matrix spectral norm.

Let $\|g\|_{L^2}^2 := \int |g(w)|^2 \mathrm{d}F_W(w)$, and consider first the L^2 rate of approximation of the series-based least squares estimator:

$$\|\widehat{\psi}_K - \psi\|_{L^2}^2 \le 2\|\widehat{\psi}_K - \psi_K\|_{L^2}^2 + 2\|\psi_K - \psi\|_{L^2}^2,$$

where it follows immediately that $\|\psi_K - \psi\|_{L^2}^2 = \xi_K^2 \le \vartheta_{0,K}^2$, implying that the best mean square approximation $\psi_K(w) = p_K(w)^\top \beta_K$ will approximate $\psi(w)$ well if $\vartheta_{0,K} \to 0$, or at least $\xi_K \to 0$, which in turn requires $K \to \infty$ in general. However, in many applications the series estimator may be misspecified or inconsistent in the sense that $\xi_K \not\to 0$. In those cases, it is natural to take $\psi_K(w)$ as the target "parameter". The following theorem establishes two distinct L^2 convergence rates for the series estimator relative to the latter quantity.

Theorem 4.1 (L_2 Convergence). *Let Assumption 4.1 hold, $\lambda_K^{-1} \zeta_{0,K}^2 \log(K)/n \to 0$, and $B_n \to 0$. Then,*

$$\|\widehat{\psi}_K - \psi_K\|_{L^2}^2 = O_P\Big(\frac{K}{n} + \mathsf{A}_n\Big),$$

where the approximation error term can be taken to be either

$$\mathsf{A}_n = \min\Big\{B_n + \xi_K^2 \,,\, R_n + \vartheta_{0,K}^2\Big\},$$

or

$$\mathsf{A}_n = B_n + B_n \min\Big\{B_n + \xi_K^2 \,,\, R_n + \vartheta_{0,K}^2\Big\} + \min\Big\{\lambda_K^{-1} \zeta_{0,K}^2 \xi_K^2/n \,,\, K\vartheta_{0,K}^2/n\Big\}.$$

8

This theorem provides new results relative to previously known mean square convergence rates for series estimation. More specifically, it allows for generated regressors based on covariates with a possibly vanishing minimum eigenvalue of the expected scaled Gram matrix (λ_K), as it may occur when the Lebesgue density of W is positive but not bounded away from zero on $\mathcal{W}$. Furthermore, the second rate estimate allows for a non-vanishing L^2 approximation error ($\vartheta_{0,K} \geq \xi_K \not\rightarrow 0$), thereby offering L^2 consistency results for general least squares approximations.

It is easy to deduce (suboptimal) uniform rates of approximation using Theorem 4.1 because

$$\max_{t \in \Lambda_q} \|\partial^t \widehat{\psi}_K - \partial^t \psi\|_\infty \leq \max_{t \in \Lambda_q} \|\partial^t p_K^\top (\widehat{\beta}_K - \beta_K)\|_\infty + \max_{t \in \Lambda_q} \|\partial^t \psi_K - \partial^t \psi\|_\infty,$$

$$\leq \zeta_{q,K} \lambda_K^{-1/2} \|\widehat{\psi}_K - \psi_K\|_{L^2} + \vartheta_{q,K},$$

where, as noted before, the first term characterizes the error in approximation when perhaps $\vartheta_{q,K} \not\rightarrow 0$, in which case the target "parameter" can be taken to be $\partial^t \psi_K(w) = \partial^t p_K(w)^\top \beta_K$, regardless of whether $K \to \infty$ or not.

Underlying the assumptions imposed in Theorem 4.1, there are several parameters that need further discussion. From the standard series estimation literature, $\zeta_{q,K} = O(K^{1+q})$ for power series and $\zeta_{q,K} = O(K^{1/2+q})$ for fourier series, splines, compact supported wavelets, and piecewise polynomial regression. Lower bounds for λ_K need to be established on a case-by-case basis when the density of W is not assumed to be bounded away from zero, so we illustrate one such verification further below for the case of rank transformations and a Gaussian copula. As already mentioned, the parameter $\xi_K \leq \vartheta_{0,K}$ captures the degree of approximation (or misspecification) of the series regression estimator, and needs not to vanish, in which case the second rate result in Theorem 4.1, and the implied uniform convergence rate, must be used. If ψ is s-times differentiable and other regularity conditions hold, then $\xi_K = O(K^{-s/d})$ for all the usual approximation basis functions if appropriately specified. In general, however, the difference between the L^2 and L^∞ approximaton errors, ξ_K and $\vartheta_{0,K}$, depends on the basis functions employed and the data generating features. In particular, for instance, when employing locally supported basis functions, it can be verified that $\xi_K \asymp \vartheta_{0,K}$, in which case $\vartheta_{q,K} = O(K^{-(s-q)/d})$ under regularity conditions. See Newey (1997), Huang (2003), Cattaneo and Farrell (2013), Belloni et al. (2015), Cattaneo et al. (2020), and references therein, for more details.

An important feature of Theorem 4.1 is that it allows for generated regressors based on the covariates, which introduces two additional quantities characterizing the approximation rate: R_n and B_n. For example, if ψ satisfies the Lipschitz condition $|\psi(a) - \psi(b)| \leq L\|a - b\|$ for some constant L, then

$$R_n = \frac{1}{n} \sum_{i=1}^n (\psi(\widehat{W}_i) - \psi(W_i))^2 \leq L^2 \max_{i=1,2,\ldots,n} \|\widehat{W}_i - W_i\|^2 = O_P(r_n),$$

and therefore the convergence rate of R_n is determined by the uniform convergence rate of the transformation of the covariates. Recall that in our application, we consider the empirical rank transformation $(W_i, \widehat{W}_i) = (U_i, \widehat{U}_i)$, and therefore $r_n = 1/n$. A similar calculation can be done to

bound B_n when $p_K(\cdot)$ is smooth, because

$$B_n \leq \lambda_K^{-1} \frac{1}{n} \sum_{i=1}^{n} \|p_K(\widehat{W}_i) - p_K(W_i)\|^2 \leq \lambda_K^{-1} L_K^2 \max_{i=1,2,\ldots,n} \|\widehat{W}_i - W_i\|^2 = O_\mathrm{P}(b_n),$$

provided that the basis functions are Lipschitz with constant L_K, and where $L_K^2 = O(\zeta_{1,K}^2)$. Thus, in the particular case of the empirical rank transformation, $b_n = \lambda_K^{-1} \zeta_{1,K}^2 / n$.

It remains to illustrate how to lower bound the minimum eigenvalue λ_K. It is well-known that if W admits a Lebesgue density bounded and bounded away from zero over the support $\mathcal{W}$, then λ_K^{-1} is uniformly bounded in K, after possibly rotating the basis functions. The following lemma considers the more interesting case when the density is not bounded away from zero, as it occurs for example when the support of W is unbounded.

Lemma 4.1 (Lower Bound on λ_K)**.** *Suppose that W admits a Lebesgue density f_W, and $p_{1K}, \ldots, p_{KK}$ are orthonormal with respect to the Lebesgue measure over the support of W. If there exists a universal constant $C > 0$ such that, for all sufficiently small $t > 0$, $\mathsf{Leb}(\{w : 0 < f_W(w) < t\}) \leq Ct^\rho$ for some $\rho > 0$, then $\lambda_K^{-1} = O(\zeta_{0,K}^{2/\rho})$.*

As expected, the additional conditions in the previous lemma restrict the tail of f_W. It remains to illustrate how to verify the result for the case when W_i and $\widehat{W}_i$ are taken to be U_i and $\widehat{U}_i$ as in Section 2. This is done in the following proposition for the case of a Gaussian copula.

Proposition 4.1 (Sufficient Conditions for Gaussian Copula)**.** *Suppose $U_1, U_2, \ldots, U_n$ follow a Gaussian copula distribution with an invertible parameter correlation matrix Σ. Then, the conditions of Lemma 4.1 are satisfied with $\rho = \lambda_{\max}^{-1}(\Sigma^{-1} - \boldsymbol{I}_d)/d$, where $\boldsymbol{I}_d$ is the d-dimensional identity matrix.*

Using Theorem 4.1 in general, or Lemma 4.1 and specific conditions on the underlying copula of the population rank-based transformed covariates, it is possible to verify the conditions of Theorem 3.1. To be more specific, Theorem 4.1 gives versatile (albeit suboptimal) uniform consistency rates for series estimators $(\widehat{\mu}_0, \widehat{\mu}_1)$ of either some approximate (misspecified) functions or the population functions (μ_0, μ_1). It is worth noting that our verification aimed for generality in terms of high-level conditions, but for the case of partition-based (locally supported) series estimators it is possible to obtain better (in fact optimal in some cases) uniform convergence rates (Cattaneo and Farrell, 2013; Belloni et al., 2015; Cattaneo et al., 2020). Specifically, Cattaneo et al. (2023) studies the case of rank-based transformations for B-splines when $d = 1$, and establishes optimal uniform convergence rates on compact support. Their results could be extended to obtain sharper uniform convergence rates with generated regressors based on the covariates as required by Theorem 3.1.

5 Generalized Framework

This section extends Rosenbaum's rank standardization idea to a more general setting, and establishes general theory that will cover Theorem 3.1 as a special case. More specifically, for $\omega = 0, 1$, consider the following general mappings

$$\phi_\omega : \mathcal{X} \to \mathcal{X}_\phi \subset \mathbb{R}^m,$$

with $\mathcal{X}$ representing the support of X and m not necessarily equal to d. Note that here we allow ϕ_0 and ϕ_1 to be different. Consider the setting when ϕ_ω is possibly unknown, and we will approximate it based on the sample $\{(X_i, D_i, Y_i)\}_{i=1}^n$, leading to a generic estimator $\widehat{\phi}_\omega$ that may differ with different ω. We then define

$$U_{\phi,\omega} := \phi_\omega(X) \qquad \text{and} \qquad \widehat{U}_{\phi,\omega,i} := \widehat{\phi}_\omega(X_i) \qquad \text{for} \quad i \in [\![n]\!].$$

Note that, when setting $\phi_0 = \phi_1 = \boldsymbol{F}$ and $\widehat{\phi}_0 = \widehat{\phi}_1 = \widehat{\boldsymbol{F}}_n$, we recover the U and $\widehat{U}_i$'s introduced in Section 2.

Similar to Section 2, let $\mathcal{J}_\phi(i)$ represent the index set of M-NN matches of $\widehat{U}_{\phi,1-D_i,i}$ in $\{\widehat{U}_{\phi,1-D_i,j} : D_j = 1 - D_i\}_{j=1}^n$ with ties broken in an arbitrary way. In other words, for determining the nearest neighbors, we are going to measure the similarity based on the Euclidean distance between transformed data points with the transformation function probably also having to be learnt from the same data. Additionally, let $\widehat{\mu}_{\phi,\omega}(u)$ be a mapping from $\mathcal{X}_\phi$ to $\mathbb{R}$ that estimates the conditional means of the outcomes

$$\mu_{\phi,\omega}(u) := \mathrm{E}[Y \,|\, U_{\phi,\omega} = u, D = \omega].$$

The general ϕ-transformation based bias-corrected matching estimator is then defined to be

$$\widehat{\tau}_\phi := \frac{1}{n} \sum_{i=1}^n \big[\widehat{Y}_{\phi,i}(1) - \widehat{Y}_{\phi,i}(0)\big],$$

where

$$\widehat{Y}_{\phi,i}(\omega) := \begin{cases} Y_i, & \text{if } D_i = \omega, \\ \frac{1}{M} \sum_{j \in \mathcal{J}_\phi(i)} (Y_j + \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) - \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,j})) & \text{if } D_i = 1 - \omega \end{cases}.$$

It follows that $\widehat{\tau}_\phi$ generalizes $\widehat{\tau}$ in (2.1).

5.1 General Theory

In order to analyze $\widehat{\tau}_\phi$, we introduce some additional notation and assumptions that are in parallel to those made in Section 3. Let the residuals from fitting the outcome models be

$$\widehat{R}_{\phi,i} := Y_i - \widehat{\mu}_{\phi,D_i}(\widehat{U}_{\phi,D_i,i}), \qquad i \in [\![n]\!],$$

and the estimator based on the outcome models be

$$\widehat{\tau}_\phi^{\mathrm{reg}} := n^{-1} \sum_{i=1}^n \big[\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i})\big].$$

Finally, let $K_\phi(i)$ be the number of matched times for the unit i according to the distances between $\widehat{U}_{\phi,D_i,i}$'s, i.e.,

$$K_\phi(i) := \sum_{j=1, D_j = 1 - D_i}^n \mathbb{1}\big(i \in \mathcal{J}_\phi(j)\big).$$

The first lemma corresponds to a generalization of the AIPW representation of the bias-corrected rank-based estimator given in (3.1) in Section 3.

11

Lemma 5.1. *It holds true that*
$$\widehat{\tau}_\phi = \widehat{\tau}_\phi^{\text{reg}} + \frac{1}{n}\sum_{i=1}^{n}(2D_i - 1)\Big(1 + \frac{K_\phi(i)}{M}\Big)\widehat{R}_{\phi,i}.$$

The first two assumptions in this section parallel Assumptions 3.1 and 3.2.

Assumption 5.1.

(i) *For almost all $x \in X$, D is independent of $(Y(0), Y(1))$ conditional on $X = x$, and there exists some constant $c > 0$ such that $c < \mathrm{P}(D = 1 \mid X = x) < 1 - c$.*

(ii) *$\{(X_i, D_i, Y_i)\}_{i=1}^{n}$ are i.i.d. following the joint distribution of (X, D, Y).*

(iii) *$\mathrm{E}[(Y(\omega) - \mu_{\phi,\omega}(U_{\phi,\omega}))^2 \mid U_{\phi,\omega} = u]$ is uniformly bounded for almost all $u \in \mathcal{X}_\phi$ and $\omega = 0, 1$.*

(iv) *$\mathrm{E}[\mu_{\phi,\omega}^2(U_{\phi,\omega})]$ is bounded for $\omega = 0, 1$.*

Assumption 5.2.

(i) *There exists a deterministic, possibly changing with n, function $\bar{\mu}_{\phi,\omega}(\cdot) : \mathbb{R}^m \to \mathbb{R}$ such that $\mathrm{E}[\bar{\mu}_{\phi,\omega}^2(U_{\phi,\omega})]$ is uniformly bounded and the estimator $\widehat{\mu}_{\phi,\omega}(x)$ satisfies $\|\widehat{\mu}_{\phi,\omega} - \bar{\mu}_{\phi,\omega}\|_\infty = o_\mathrm{P}(1)$ for $\omega = 0, 1$.*

(ii) *$\max_{i \in [\![n]\!]} |\bar{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) - \bar{\mu}_{\phi,\omega}(U_{\phi,\omega,i})| = o_\mathrm{P}(1)$ for $\omega = 0, 1$.*

The next assumption regulates the transformation ϕ; cf. Lin et al. (2023, Section B) and Lin and Han (2022, Assumption 3.3(ii)). From a high level perspective, it roughly states that $M^{-1}K_\phi(i)$ should be a consistent density ratio estimator. A detailed discussion of this assumption is given in Section 5.2 ahead.

Assumption 5.3. *The number of matched times satisfies*
$$\lim_{n\to\infty} \mathrm{E}\Big[\frac{K_\phi(1)}{M} - \Big(D_1\frac{1 - e(X_1)}{e(X_1)} + (1 - D_1)\frac{e(X_1)}{1 - e(X_1)}\Big)\Big]^2 = 0,$$
where, for any $x \in \mathcal{X}$, $e(x) := \mathrm{P}(D = 1 \mid X = x)$ is the propensity score.

The next three assumptions correspond to Assumptions 3.4 through 3.6 in Section 3.

Assumption 5.4.

(i) *The estimator $\widehat{\mu}_{\phi,\omega}(x)$ satisfies $\|\widehat{\mu}_{\phi,\omega} - \mu_{\phi,\omega}\|_\infty = o_\mathrm{P}(1)$ for $\omega = 0, 1$.*

(ii) *$\max_{i \in [\![n]\!]} |\mu_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) - \mu_{\phi,\omega}(U_{\phi,\omega,i})| = o_\mathrm{P}(1)$ for $\omega = 0, 1$.*

(iii) *$\mathrm{E}[Y(\omega) \mid X = x] = \mu_{\phi,\omega}(\phi_\omega(x))$ for almost all $x \in \mathcal{X}$ and $\omega = 0, 1$.*

Assumption 5.5.

(i) *$\mathrm{E}[(Y(\omega) - \mu_{\phi,\omega}(U_{\phi,\omega}))^2 \mid U_{\phi,\omega} = u]$ is uniformly bounded away from zero for almost all $u \in \mathcal{X}_\phi$ and $\omega = 0, 1$.*

(ii) *There exists a constant $c > 0$ such that $\mathrm{E}[|Y(\omega) - \mu_{\phi,\omega}(U_{\phi,\omega})|^{2+c} \mid U_{\phi,\omega} = u]$ is uniformly bounded for almost all $u \in \mathcal{X}_\phi$ and $\omega = 0, 1$.*

12

(iii) $\max_{t\in\Lambda_{\max\{\lfloor m/2\rfloor,1\}+1}}\|\partial^t\mu_{\phi,\omega}\|_\infty$ *is bounded.*

(iv) $\mathrm{E}[Y(\omega)\,|\,X=x]=\mu_{\phi,\omega}(\phi_\omega(x))$ *for almost all $x\in\mathcal{X}$ and $\omega=0,1$.*

(v) *The density of $\phi_\omega(X)$ is continuous over its support for $\omega=0,1$.*

Assumption 5.6. *For $\omega=0,1$, the estimator $\widehat{\mu}_\omega(x)$ satisfies*

$$\max_{t\in\Lambda_{\max\{\lfloor m/2\rfloor,1\}+1}}\|\partial^t\widehat{\mu}_{\phi,\omega}\|_\infty=O_{\mathrm{P}}(1)\ \ \text{and}\ \ \max_{t\in\Lambda_\ell}\|\partial^t\widehat{\mu}_{\phi,\omega}-\partial^t\mu_{\phi,\omega}\|_\infty=O_{\mathrm{P}}(n^{-\gamma_\ell})\ \text{for all }\ell\in[\![\max\{\lfloor m/2\rfloor,1\}]\!],$$

with some constants $\gamma_\ell>\max\{\frac{1}{2}-\frac{\ell}{m},0\}$ for $\ell=1,2,\ldots,\max\{\lfloor m/2\rfloor,1\}$.

The next assumption poses a Donsker-type condition on the approximation accuracy of the estimated transformation $\widehat{\phi}_\omega$ towards ϕ_ω. This assumption is usually needed when one wishes to avoid using sample splitting.

Assumption 5.7. *For $\omega=0,1$, the estimator $\widehat{\phi}_\omega$ satisfies*

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathrm{P}\Big(\sqrt{n}\sup_{x,y\in\mathcal{X},\|\phi_\omega(x)-\phi_\omega(y)\|\le\delta}\|(\widehat{\phi}_\omega-\phi_\omega)(x)-(\widehat{\phi}_\omega-\phi_\omega)(y)\|\ge\epsilon\Big)=0.$$

We are now ready to present the following theorem, which is a generalization to Theorem 3.1.

Theorem 5.1 (Generalized Main Theorem).

(i) *(Double robustness of $\widehat{\tau}_\phi$) If either Assumptions 5.1, 5.2, 5.3 hold, or Assumptions 5.1 and 5.4 hold, then*

$$\widehat{\tau}_\phi-\tau\xrightarrow{\mathsf{p}}0.$$

(ii) *(Semiparametric efficiency of $\widehat{\tau}_\phi$) Assume the distribution of (X,D,Y) satisfies Assumptions 5.1, 5.3, 5.5, 5.6, 5.7. Define*

$$\gamma=\max\left\{\Big[1-\frac{1}{2}\frac{m}{\max\{\lfloor m/2\rfloor,1\}+1}\Big],\min_{\ell\in[\![\max\{\lfloor m/2\rfloor,1\}]\!]}\Big[1-\Big(\frac{1}{2}-\gamma_\ell\Big)\frac{m}{\ell}\Big]\right\},$$

recalling that γ_ℓ's were introduced in Assumption 5.6. Then, if $M\to\infty$ and $M/n^\gamma\to 0$ as $n\to\infty$, we have

$$\sqrt{n}(\widehat{\tau}_\phi-\tau)\xrightarrow{\mathsf{d}}N(0,\sigma^2).$$

(iii) *If in addition Assumption 5.4 holds, then $\widehat{\sigma}^2_\phi\xrightarrow{\mathsf{p}}\sigma^2$, where*

$$\widehat{\sigma}^2_\phi:=\frac{1}{n}\sum_{i=1}^n\Big[\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i})-\widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i})+(2D_i-1)\Big(1+\frac{K_\phi(i)}{M}\Big)\widehat{R}_{\phi,i}-\widehat{\tau}_\phi\Big]^2.$$

5.2 Discussion on High-Level Assumption

It remains to decipher the high-level condition in Assumption 5.3. To this end, we first give additional regularizations about the population-transformed data.

Assumption 5.8.

(i) The diameter of $\mathcal{X}_\phi$ and the surface area of $\mathcal{X}_\phi$ are bounded.

(ii) The density of $\phi_\omega(X)$ is continuous over its support for $\omega = 0, 1$.

(iii) $\mathrm{P}(D = 1 \,|\, \phi_\omega(X) = \phi_\omega(x)) = \mathrm{P}(D = 1 \,|\, X = x)$ for almost all $x \in \mathcal{X}$ and $\omega = 0, 1$.

Next, we give two different type of conditions for the estimator $\widehat{\phi}_\omega$ to approximate ϕ_ω so that Assumption 5.3 can hold.

Assumption 5.9. *For $\omega = 0, 1$,*

$$\lim_{n \to \infty} \mathrm{E}\Big[\Big(\frac{n}{M}\Big)^2 \sup_{x_1, x_2 \in \mathcal{X}} \|\widehat{\phi}_\omega(\cdot; x_1, x_2) - \phi_\omega\|_\infty^{2d}\Big] = 0,$$

where $\widehat{\phi}_\omega(\cdot; x_1, x_2)$ is the estimator constructed by inserting two more new points, x_1 and x_2, into the group with $D = 1 - \omega$ for some $x_1, x_2 \in \mathcal{X}$.

Assumption 5.10. *For $\omega = 0, 1$, we assume that for any fixed $\epsilon > 0$, there exists a function $T_\epsilon(u)$ such that, for any $\delta > 0$,*

$$\mathrm{P}\Big(\sup_{\delta \geq u} \delta^{-1} \sup_{\|\phi_\omega(s) - \phi_\omega(t)\| \leq \delta} \sup_{x_1, x_2 \in \mathcal{X}} \|(\widehat{\phi}_\omega(\cdot; x_1, x_2) - \phi_\omega)(s) - (\widehat{\phi}_\omega(\cdot; x_1, x_2) - \phi_\omega)(t)\| > \epsilon\Big) \leq T_\epsilon(u),$$

and, for any $k \in [\![2]\!]$,

$$\lim_{n \to \infty} \Big(\frac{n}{M}\Big)^k \int_0^\infty u^{k-1} T_\epsilon(u^{1/m}) \mathrm{d}u = 0$$

and

$$\lim_{n \to \infty} \Big(\frac{n}{M}\Big)^2 \mathrm{P}\Big(\|\widehat{\phi}_\omega - \phi_\omega\|_\infty > \epsilon\Big) = 0.$$

Theorem 5.2. *Assume that Assumption 5.8 holds, $M \log n / n \to 0$ and $M \to \infty$ as $n \to \infty$. If either Assumption 5.9 or Assumption 5.10 holds, then Assumption 5.3 holds.*

6 Conclusion

This paper studied the large sample properties of Rosenbaum's rank-based matching estimator with regression adjustment, and established its consistency, doubly robutness, asymptotic normality, and semiparametric efficiency. Consistency of a plug-in variance estimator was also established. These results were obtained as a consequence of a more general theorem that allows for a class of transformations of the covariates, an leading special case being the empirical rank transformation proposed by Rosenbaum (2005, 2020). To provide primitive conditions for regression adjustment, novel convergence rates for series estimators with generated regressors and possibly covariate Lebesgue density not bounded away from zero were derived, which may be of independent interest.

Acknowledgments

A   Proofs of Main Results

A.1   Proof of Theorem 3.1

We take $\phi_0 = \phi_1 = \boldsymbol{F}$ and $\widehat{\phi}_0 = \widehat{\phi}_1 = \widehat{\boldsymbol{F}}_n$. Note that $\boldsymbol{F}$ is a bijective function. Then Assumption 3.1 implies Assumption 5.1. To show that Assumption 3.2 and Assumption 3.4 imply Assumption 5.2 and Assumption 5.4, respectively, it remains to show

$$\max_{i \in [\![n]\!]} |\bar{\mu}_\omega(\widehat{U}_i) - \bar{\mu}_\omega(U_i)| = o_{\mathrm{P}}(1) \quad \text{or} \quad \max_{i \in [\![n]\!]} |\mu_\omega(\widehat{U}_i) - \mu_\omega(U_i)| = o_{\mathrm{P}}(1).$$

Note that $[0,1]^d$ is compact. Then the continuity of μ_ω implies uniform continuity. Then $\max_{i \in [\![n]\!]} |\mu_\omega(\widehat{U}_i) - \mu_\omega(U_i)| = o_{\mathrm{P}}(1)$ is directly from $\max_{i \in [\![n]\!]} \|\widehat{U}_i - U_i\| = o_{\mathrm{P}}(1)$. The same holds for $\bar{\mu}_\omega$.

To verify Assumption 5.3, we use Theorem 5.2. Assumption 5.8 holds by Assumption 3.3 and the bijection of $\boldsymbol{F}$. We verify Assumption 5.9 when $d \geq 2$ and Assumption 5.10 when $d = 1$.

Part I. $d \geq 2$.

Note that

$$\sup_{x_1, x_2 \in \mathcal{X}} \|\widehat{\boldsymbol{F}}_n(\cdot; x_1, x_2) - \boldsymbol{F}\|_\infty^{2d} \lesssim \sup_{x_1, x_2 \in \mathcal{X}} \|\widehat{\boldsymbol{F}}_n(\cdot; x_1, x_2) - \widehat{\boldsymbol{F}}_n\|_\infty^{2d} + \|\widehat{\boldsymbol{F}}_n - \boldsymbol{F}\|_\infty^{2d},$$

where $\lesssim$ means "asymptotically small than".

By the definition of $\widehat{\boldsymbol{F}}_n(\cdot; x_1, x_2)$ and $\widehat{\boldsymbol{F}}_n$, for any $x_1, x_2 \in \mathcal{X}$,

$$\|\widehat{\boldsymbol{F}}_n(\cdot; x_1, x_2) - \widehat{\boldsymbol{F}}_n\|_\infty$$

$$\leq \sqrt{d} \max_{k \in [\![d]\!]} \max_{x \in \mathbb{R}} \left| \frac{1}{n+2} \Big(\sum_{i=1}^n \mathbb{1}(X_{i,k} \leq x) + \mathbb{1}(x_1 \leq x) + \mathbb{1}(x_2 \leq x) \Big) - \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_{i,k} \leq x) \right|$$

$$= \sqrt{d} \max_{k \in [\![d]\!]} \max_{x \in \mathbb{R}} \left| \frac{1}{n+2} \Big(\mathbb{1}(x_1 \leq x) + \mathbb{1}(x_2 \leq x) \Big) - \frac{2}{n(n+2)} \sum_{i=1}^n \mathbb{1}(X_{i,k} \leq x) \right|$$

$$\leq \frac{4\sqrt{d}}{n+2}.$$

We then have

$$\lim_{n \to \infty} \mathrm{E}\left[\Big(\frac{n}{M} \Big)^2 \sup_{x_1, x_2 \in \mathcal{X}} \|\widehat{\boldsymbol{F}}_n(\cdot; x_1, x_2) - \widehat{\boldsymbol{F}}_n\|_\infty^{2d} \right] = 0.$$

Note that

$$\|\widehat{\boldsymbol{F}}_n - \boldsymbol{F}\|_\infty^{2d} \lesssim \max_{k \in [\![d]\!]} \max_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_{i,k} \leq x) - \mathrm{P}(X_k \leq x) \right|^{2d}$$

$$\leq \sum_{k=1}^d \max_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_{i,k} \leq x) - \mathrm{P}(X_k \leq x) \right|^{2d}.$$

15

By the Dvoretzky–Kiefer–Wolfowitz inequality, we have

$$\mathrm{E}[\|\widehat{\boldsymbol{F}}_n - \boldsymbol{F}\|_\infty^{2d}] \lesssim n^{-d}.$$

By $d \geq 2$ and $M \to \infty$, we have

$$\lim_{n \to \infty} \mathrm{E}\Big[\Big(\frac{n}{M}\Big)^2 \|\widehat{\boldsymbol{F}}_n - \boldsymbol{F}\|_\infty^{2d}\Big] = 0.$$

The proof is now complete.

Part II. $d = 1$.

Note that for any $s, t \in \mathbb{R}$, by Part I,

$$\sup_{x_1, x_2 \in \mathcal{X}} \Big|(\widehat{\boldsymbol{F}}_n(\cdot; x_1, x_2) - \boldsymbol{F})(s) - (\widehat{\boldsymbol{F}}_n(\cdot; x_1, x_2) - \boldsymbol{F})(t)\Big| \leq \frac{8\sqrt{d}}{n+2} + \Big|(\widehat{\boldsymbol{F}}_n - \boldsymbol{F})(s) - (\widehat{\boldsymbol{F}}_n - \boldsymbol{F})(t)\Big|.$$

For any $\epsilon > 0$, we can take n sufficiently large such that $\frac{8\sqrt{d}}{n+2} < \epsilon$, and then

$$\mathrm{P}\Big(\sup_{\delta \geq u} \delta^{-1} \sup_{|\boldsymbol{F}(s) - \boldsymbol{F}(t)| \leq \delta} \sup_{x_1, x_2 \in \mathcal{X}} \Big|(\widehat{\boldsymbol{F}}_n(\cdot; x_1, x_2) - \boldsymbol{F})(s) - (\widehat{\boldsymbol{F}}_n(\cdot; x_1, x_2) - \boldsymbol{F})(t)\Big| > 2\epsilon\Big)$$

$$\leq \mathrm{P}\Big(\sup_{\delta \geq u} \delta^{-1} \sup_{|\boldsymbol{F}(s) - \boldsymbol{F}(t)| \leq \delta} \Big|(\widehat{\boldsymbol{F}}_n - \boldsymbol{F})(s) - (\widehat{\boldsymbol{F}}_n - \boldsymbol{F})(t)\Big| > \epsilon\Big)$$

$$= \mathrm{P}\Big(\sup_{\delta \geq u} \sup_{s \leq t: \mathrm{P}(s < X \leq t) \leq \delta} \delta^{-1} \Big|\frac{1}{n} \sum_{i=1}^n \mathbb{1}(s < X_i \leq t) - \mathrm{P}(s < X \leq t)\Big| > \epsilon\Big).$$

Fix $u \geq 0$ and ϵ. Let $T_i = \delta^{-1} u(\mathbb{1}(s < X_i \leq t) - \mathrm{P}(s < X \leq t))$ for $i \in [\![n]\!]$. Then for any $i \in [\![n]\!]$, we have $\mathrm{E}[T_i] = 0$, and that $|T_i| \leq 1$ for $\delta \geq u$.

Note that

$$\sup_{\delta \geq u} \sup_{s \leq t: \mathrm{P}(s < X \leq t) \leq \delta} \sum_{i=1}^n \mathrm{E}[T_i^2] \leq \sup_{\delta \geq u} \sup_{s \leq t: \mathrm{P}(s < X \leq t) \leq \delta} \sum_{i=1}^n \delta^{-2} u^2 \mathrm{P}(s < X_i \leq t) \leq un,$$

and

$$\mathrm{E}[\sup_{\delta \geq u} \sup_{s \leq t: \mathrm{P}(s < X \leq t) \leq \delta} \sum_{i=1}^n T_i^2] \leq \mathrm{E}[\sup_{\delta \geq u} \sup_{s \leq t: \mathrm{P}(s < X \leq t) \leq \delta} \sum_{i=1}^n \delta^{-2} u^2 \mathbb{1}(s < X_i \leq t)] \lesssim un,$$

by standard empirical process theory. Then by the concentration inequality for bounded processes (Boucheron et al., 2013, Theorem 12.2), we have for n sufficiently large,

$$\mathrm{P}\Big(\sup_{\delta \geq u} \sup_{s \leq t: \mathrm{P}(s < X \leq t) \leq \delta} \delta^{-1} \Big|\frac{1}{n} \sum_{i=1}^n \mathbb{1}(s < X_i \leq t) - \mathrm{P}(s < X \leq t)\Big| > \epsilon\Big)$$

$$= \mathrm{P}\Big(\sup_{\delta \geq u} \sup_{s \leq t: \mathrm{P}(s < X \leq t) \leq \delta} \Big|\sum_{i=1}^n T_i\Big| > un\epsilon\Big)$$

$$\leq \exp\Big(-\frac{u^2 n^2 \epsilon^2}{Cun + un\epsilon}\Big) = \exp\Big(-\frac{\epsilon^2}{C + \epsilon} un\Big).$$

The proof for this part is now complete by taking integral using $m = 1$ and $M \to \infty$.

Lastly, it is easy to see Assumption 3.5 and Assumption 3.6 imply Assumption 5.5 and Assumption 5.6, respectively. Assumption 5.7 is followed by the Donsker's theorem applied to empirical distribution function.

A.2   Proof of Theorem 4.1

Let $\widehat{\boldsymbol{Q}} := \boldsymbol{P}^\top \boldsymbol{P}/n$. Using $\|\boldsymbol{Q}^{-1/2}\|_2 = \lambda_K^{-1/2}$,

$$\|\boldsymbol{Q}^{-1/2}p_K(W_i)p_K(W_i)^\top\boldsymbol{Q}^{-1/2}\|_2 = \|\boldsymbol{Q}^{-1/2}p_K(W_i)\|^2 \le \lambda_K^{-1}\zeta_{0,K}^2,$$

$$\|\mathrm{E}[\boldsymbol{Q}^{-1/2}p_K(W_i)p_K(W_i)^\top\boldsymbol{Q}^{-1}p_K(W_i)p_K(W_i)^\top\boldsymbol{Q}^{-1/2}]\|_2^2$$
$$\le \lambda_K^{-1}\zeta_{0,K}^2\|\mathrm{E}[\boldsymbol{Q}^{-1/2}p_K(W_i)p_K(W_i)^\top\boldsymbol{Q}^{-1/2}]\|_2^2 = \lambda_K^{-1}\zeta_{0,K}^2,$$

and a standard exponential concentration inequality for random matrices (Tropp, 2012, Section 6),

$$\|\boldsymbol{Q}^{-1/2}\widehat{\boldsymbol{Q}}\boldsymbol{Q}^{-1/2} - \boldsymbol{I}_K\|_2 = O_\mathrm{P}\big(\lambda_K^{-1/2}\zeta_{0,K}\sqrt{\log(K)/n} + \lambda_K^{-1}\zeta_{0,K}^2\log(K)/n\big) = o_\mathrm{P}(1)$$

because $\lambda_K^{-1}\zeta_{0,K}^2\log(K)/n \to 0$ by assumption.

Let $\widehat{\boldsymbol{Q}}_n := \boldsymbol{P}_n^\top \boldsymbol{P}_n/n$. Then,

$$\|\boldsymbol{Q}^{-1/2}\boldsymbol{P}^\top\|_2^2 = \|\boldsymbol{Q}^{-1/2}\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{Q}^{-1/2}\|_2$$
$$\le n\|\boldsymbol{Q}^{-1/2}(\widehat{\boldsymbol{Q}} - \boldsymbol{Q})\boldsymbol{Q}^{-1/2}\|_2 + n\|\boldsymbol{Q}^{-1/2}\boldsymbol{Q}\boldsymbol{Q}^{-1/2}\|_2$$
$$= O_\mathrm{P}(n\lambda_K^{-1/2}\zeta_{0,K}\sqrt{\log(K)/n} + n),$$

and therefore

$$\|\boldsymbol{Q}^{-1/2}(\widehat{\boldsymbol{Q}}_n - \widehat{\boldsymbol{Q}})\boldsymbol{Q}^{-1/2}\|_2$$
$$= \|\boldsymbol{Q}^{-1/2}(\boldsymbol{P}_n^\top\boldsymbol{P}_n - \boldsymbol{P}^\top\boldsymbol{P})\boldsymbol{Q}^{-1/2}\|_2/n$$
$$\le \|\boldsymbol{Q}^{-1/2}(\boldsymbol{P}_n - \boldsymbol{P})^\top(\boldsymbol{P}_n - \boldsymbol{P})\boldsymbol{Q}^{-1/2}\|_2/n + 2\|\boldsymbol{Q}^{-1/2}\boldsymbol{P}^\top(\boldsymbol{P}_n - \boldsymbol{P})\boldsymbol{Q}^{-1/2}\|_2/n$$
$$\le \|(\boldsymbol{P}_n - \boldsymbol{P})\boldsymbol{Q}^{-1/2}\|_2^2/n + 2\|\boldsymbol{Q}^{-1/2}\boldsymbol{P}^\top\|_2\|(\boldsymbol{P}_n - \boldsymbol{P})\boldsymbol{Q}^{-1/2}\|_2/n$$
$$= O_\mathrm{P}(B_n + \lambda_K^{-1/4}\zeta_{0,K}(\log(K)/n)^{1/4}B_n^{1/2} + B_n^{1/2}) = o_\mathrm{P}(1),$$

because $\lambda_K^{-1}\zeta_{0,K}^2\log(K)/n \to 0$ and $B_n \to 0$ by assumption.

Putting the two results together,

$$\|\boldsymbol{Q}^{-1/2}\widehat{\boldsymbol{Q}}_n\boldsymbol{Q}^{-1/2} - \boldsymbol{I}_K\|_2 \le \|\boldsymbol{Q}^{-1/2}(\widehat{\boldsymbol{Q}}_n - \widehat{\boldsymbol{Q}})\boldsymbol{Q}^{-1/2}\|_2 + \|\boldsymbol{Q}^{-1/2}\widehat{\boldsymbol{Q}}\boldsymbol{Q}^{-1/2} - \boldsymbol{I}_K\|_2$$
$$= O_\mathrm{P}(B_n^{1/2} + \lambda_K^{-1/2}\zeta_{0,K}\sqrt{\log(K)/n}) = o_\mathrm{P}(1),$$

given the rate restrictions imposed in the theorem.

Let $\mathbb{1}_n = \mathbb{1}(\lambda_{\min}(\boldsymbol{Q}^{-1/2}\widehat{\boldsymbol{Q}}_n\boldsymbol{Q}^{-1/2}) > 1/2)$. Then, $\lim_{n\to\infty}\mathrm{P}(\mathbb{1}_n = 1) = 1$. Letting $\varepsilon := \boldsymbol{Y} - \boldsymbol{\Psi}$,

$$\mathbb{1}_n\|\widehat{\psi}_K - \psi_K\|_{L^2}^2 = \mathbb{1}_n\int[p_K(w)^\top\widehat{\beta}_K - p_K(w)^\top\beta_K]^2 dF_W(w)$$
$$= \mathbb{1}_n(\widehat{\beta}_K - \beta_K)^\top\mathrm{E}[p_K(W)p_K(W)^\top](\widehat{\beta}_K - \beta_K) = \mathbb{1}_n\|\boldsymbol{Q}^{1/2}(\widehat{\beta}_K - \beta_K)\|^2$$
$$\le 2\mathbb{1}_n\|\boldsymbol{Q}^{1/2}\widehat{\boldsymbol{Q}}_n^{-1}\boldsymbol{P}_n^\top\varepsilon/n\|^2 + 2\mathbb{1}_n\|\boldsymbol{Q}^{1/2}\widehat{\boldsymbol{Q}}_n^{-1}\boldsymbol{P}_n^\top(\boldsymbol{\Psi} - \boldsymbol{P}_n\beta_K)/n\|^2.$$

For the first term, we have

$$\mathbb{1}_n\|\boldsymbol{Q}^{1/2}\widehat{\boldsymbol{Q}}_n^{-1}\boldsymbol{P}_n^\top\varepsilon/n\|^2 = O_\mathrm{P}(K/n)$$

because

$$\mathbb{1}_n\|\boldsymbol{Q}^{1/2}\widehat{\boldsymbol{Q}}_n^{-1}\boldsymbol{P}_n^\top\varepsilon/n\|^2 \le \mathbb{1}_n\|\boldsymbol{Q}^{1/2}\widehat{\boldsymbol{Q}}_n^{-1/2}\|_2^2\|\widehat{\boldsymbol{Q}}_n^{-1/2}\boldsymbol{P}_n^\top\varepsilon/n\|^2 = O_\mathrm{P}(1)\mathbb{1}_n\|\widehat{\boldsymbol{Q}}_n^{-1/2}\boldsymbol{P}_n^\top\varepsilon/n\|^2,$$

17

using $\mathbb{1}_n\|\boldsymbol{Q}^{1/2}\widehat{\boldsymbol{Q}}_n^{-1/2}\|_2^2 = \mathbb{1}_n\|\boldsymbol{Q}^{1/2}\widehat{\boldsymbol{Q}}_n^{-1}\boldsymbol{Q}^{1/2}\|_2 = O_{\mathrm{P}}(1)$, and $\mathrm{E}[\|\widehat{\boldsymbol{Q}}_n^{-1/2}\boldsymbol{P}_n^\top\varepsilon/n\|^2|W_1,\ldots,W_n] = O_{\mathrm{P}}(K/n)$ since $\mathrm{Var}(Y|W=w)$ is uniformly bounded.

We can bound the second term in different ways, depending on the approximation errors considered. The first two bounds rely on vanishing approximation errors ($\xi_K \to 0$ or $\vartheta_{0,K} \to 0$), and thus (implicitly) require $K \to \infty$ in general:

$$\mathbb{1}_n\|\boldsymbol{Q}^{1/2}\widehat{\boldsymbol{Q}}_n^{-1}\boldsymbol{P}_n^\top(\boldsymbol{\Psi} - \boldsymbol{P}_n\beta_K)/n\|^2$$
$$\leq \mathbb{1}_n\|\boldsymbol{Q}^{1/2}\widehat{\boldsymbol{Q}}_n^{-1}\boldsymbol{P}_n^\top/\sqrt{n}\|_2^2\|(\boldsymbol{\Psi} - \boldsymbol{P}_n\beta_K)/\sqrt{n}\|^2$$
$$\leq O_{\mathrm{P}}(1)\|\boldsymbol{\Psi} - \boldsymbol{P}_n\beta_K\|^2/n$$
$$= O_{\mathrm{P}}(\min\{B_n + \xi_K^2, R_n + \vartheta_{0,K}^2\}),$$

because $\mathbb{1}_n\|\boldsymbol{Q}^{1/2}\widehat{\boldsymbol{Q}}_n^{-1}\boldsymbol{P}_n^\top/\sqrt{n}\|_2^2 = \mathbb{1}_n\|\boldsymbol{Q}^{1/2}\widehat{\boldsymbol{Q}}_n^{-1}\boldsymbol{Q}^{1/2}\|_2 = O_{\mathrm{P}}(1)$, and because the term $\|\boldsymbol{\Psi} - \boldsymbol{P}_n\beta_K\|^2/n$ can be bounded in two different ways:

$$\|\boldsymbol{\Psi} - \boldsymbol{P}_n\beta_K\|^2/n \leq 2\|\boldsymbol{\Psi} - \boldsymbol{\Psi}_n\|^2/n + 2\|\boldsymbol{\Psi}_n - \boldsymbol{P}_n\beta_K\|^2/n = O_{\mathrm{P}}(R_n + \vartheta_{0,K}^2),$$

or

$$\|\boldsymbol{\Psi} - \boldsymbol{P}_n\beta_K\|^2/n \leq 2\|\boldsymbol{\Psi} - \boldsymbol{P}\beta_K\|^2/n + 2\|(\boldsymbol{P} - \boldsymbol{P}_n)\beta_K\|^2/n$$
$$\leq O_{\mathrm{P}}(\xi_K^2) + 2\|(\boldsymbol{P} - \boldsymbol{P}_n)\boldsymbol{Q}^{-1/2}\|_2^2\|\mathrm{E}[\boldsymbol{Q}^{-1/2}p_K(W_1)\psi(W_1)]\|^2/n$$
$$= O_{\mathrm{P}}(B_n + \xi_K^2),$$

because $\|\mathrm{E}[\boldsymbol{Q}^{-1/2}p_K(W_i)\psi(W_i)] = \beta_K^\top\boldsymbol{Q}\beta_K = \mathrm{E}[(p_K(W_i)^\top\beta_K)^2] = \mathrm{E}[\psi_K(W_i)^2] \leq \mathrm{E}[\psi(W_i)^2] = O(1)$. Therefore, $\|(\boldsymbol{P} - \boldsymbol{P}_n)\beta_K\|^2/n = O_{\mathrm{P}}(B_n)$.

Next, for other possible bounds that do not require vanishing approximation errors ($\vartheta_{0,K} \geq \xi_K \not\to 0$), even when possibly $K \to \infty$, notice that

$$\mathbb{1}_n\|\boldsymbol{Q}^{1/2}\widehat{\boldsymbol{Q}}_n^{-1}\boldsymbol{P}_n^\top(\boldsymbol{\Psi} - \boldsymbol{P}_n\beta_K)/n\|^2$$
$$\leq \mathbb{1}_n\|\boldsymbol{Q}^{1/2}\widehat{\boldsymbol{Q}}_n^{-1}\boldsymbol{Q}^{1/2}\|_2^2\|\boldsymbol{Q}^{-1/2}\boldsymbol{P}_n^\top(\boldsymbol{\Psi} - \boldsymbol{P}_n\beta_K)\|^2/n^2$$
$$\leq O_{\mathrm{P}}(1)\|\boldsymbol{Q}^{-1/2}(\boldsymbol{P}_n - \boldsymbol{P})^\top(\boldsymbol{\Psi} - \boldsymbol{P}_n\beta_K)\|^2/n^2$$
$$\quad + O_{\mathrm{P}}(1)\|\boldsymbol{Q}^{-1/2}\boldsymbol{P}^\top(\boldsymbol{P}_n - \boldsymbol{P})\beta_K\|^2/n^2$$
$$\quad + O_{\mathrm{P}}(1)\|\boldsymbol{Q}^{-1/2}\boldsymbol{P}^\top(\boldsymbol{\Psi} - \boldsymbol{P}\beta_K)\|^2/n^2,$$

where each of the three terms are bounded as follows. For the first term,

$$\|\boldsymbol{Q}^{-1/2}(\boldsymbol{P}_n - \boldsymbol{P})^\top(\boldsymbol{\Psi} - \boldsymbol{P}_n\beta_K)\|^2/n^2$$
$$\leq \|\boldsymbol{Q}^{-1/2}(\boldsymbol{P}_n - \boldsymbol{P})^\top\|_2^2\|\boldsymbol{\Psi} - \boldsymbol{P}_n\beta_K\|^2/n^2$$
$$= O_{\mathrm{P}}(B_n)O_{\mathrm{P}}(\min\{B_n + \xi_K^2, R_n + \vartheta_{0,K}^2\}),$$

using the calcultions above. For the second term,

$$\|\boldsymbol{Q}^{-1/2}\boldsymbol{P}^\top(\boldsymbol{P}_n - \boldsymbol{P})\beta_K\|^2/n^2$$
$$\leq \|\boldsymbol{Q}^{-1/2}\boldsymbol{P}^\top\|_2^2\|(\boldsymbol{P}_n - \boldsymbol{P})\beta_K\|^2/n^2$$
$$= O_{\mathrm{P}}(\lambda_K^{-1/2}\zeta_{0,K}\sqrt{\log(K)/n} + 1)O_{\mathrm{P}}(B_n),$$

18

also using the calculations above. Finally, for the third and the last term,

$$\|\boldsymbol{Q}^{-1/2}\boldsymbol{P}^\top(\boldsymbol{\Psi} - \boldsymbol{P}\beta_K)\|^2/n^2 = O_\mathrm{P}(\min\{\lambda_K^{-1}\zeta_{0,K}^2\xi_K^2, K\vartheta_{0,K}^2\}/n)$$

because, by the orthogonality of the L^2 projection,

$$\mathrm{E}[\|\boldsymbol{Q}^{-1/2}\boldsymbol{P}^\top(\boldsymbol{\Psi} - \boldsymbol{P}\beta_K)\|^2]/n^2$$

$$= \frac{1}{n^2}\mathrm{E}\Big[\Big(\sum_{i=1}^n \boldsymbol{Q}^{-1/2}p_K(W_i)(\psi(W_i) - \psi_K(W_i))\Big)^\top\Big(\sum_{i=1}^n \boldsymbol{Q}^{-1/2}p_K(W_i)(\psi(W_i) - \psi_K(W_i))\Big)\Big]$$

$$= \frac{1}{n}\mathrm{E}\Big[p_K(W_i)^\top\boldsymbol{Q}^{-1}p_K(W_i)(\psi(W_i) - \psi_K(W_i))^2\Big].$$

The final result in the theorem follows because $\lim_{n\to\infty}\mathrm{P}(\mathbb{1}_n = 1) = 1$.

A.3   Proof of Lemma 4.1

For any $u \in \mathbb{R}^K$ with $\|u\| = 1$, by the orthonormality of the basis functions, we have

$$1 = \|u\|^2 = \int \Big(\sum_{k=1}^K u_k p_{kK}(w)\Big)^2 \mathrm{d}w.$$

Note that

$$\lambda_{\min}(\mathrm{E}[p_K(W_1)p_K(W_1)^\top]) = \min_{u\in\mathbb{R}^K:\|u\|=1} u^\top\mathrm{E}[p_K(W_1)p_K(W_1)^\top]u = \min_{u\in\mathbb{R}^K:\|u\|=1} \mathrm{E}\Big[\Big(\sum_{k=1}^K u_k p_{kK}(W_1)\Big)^2\Big]$$

$$= \min_{u\in\mathbb{R}^K:\|u\|=1} \int \Big(\sum_{k=1}^K u_k p_{kK}(w)\Big)^2 f_W(w)\mathrm{d}w,$$

and since $\mathrm{E}[p_K(W_1)p_K(W_1)^\top]$ is positive semidefinite,

$$\|\mathrm{E}[p_K(W_1)p_K(W_1)^\top]\|_2 = \max_{u\in\mathbb{R}^K:\|u\|=1} \int \Big(\sum_{k=1}^K u_k p_{kK}(w)\Big)^2 f_W(w)\mathrm{d}w.$$

If f_W is bounded away from zero over the support of W, then for any $u \in \mathbb{R}^K$ with $\|u\| = 1$,

$$\int \Big(\sum_{k=1}^K u_k p_{kK}(w)\Big)^2 f_W(w)\mathrm{d}w \geq c\int \Big(\sum_{k=1}^K u_k p_{kK}(w)\Big)^2 \mathrm{d}w = c,$$

for some constants $c > 0$. If f_W is bounded over the support of W, then

$$\int \Big(\sum_{k=1}^K u_k p_{kK}(w)\Big)^2 f_W(w)\mathrm{d}w \leq C\int \Big(\sum_{k=1}^K u_k p_{kK}(w)\Big)^2 \mathrm{d}w = C,$$

for some constants $C > 0$.

More generally, for any $t > 0$,

$$\int \Big(\sum_{k=1}^K u_k p_{kK}(w)\Big)^2 f_W(w)\mathrm{d}w \geq \int \Big(\sum_{k=1}^K u_k p_{kK}(w)\Big)^2 f_W(w)\mathbb{1}(f_W(w) \geq t)\mathrm{d}w$$

$$\geq t \int \Big(\sum_{k=1}^{K} u_k p_{kK}(w) \Big)^2 \mathbb{1}(f_W(w) \geq t) \mathrm{d}w$$

$$= t \Big[1 - \int \Big(\sum_{k=1}^{K} u_k p_{kK}(w) \Big)^2 \mathbb{1}(0 < f_W(w) < t) \mathrm{d}w \Big].$$

By the Cauchy-Schwarz inequality, for all sufficiently small $t > 0$,

$$\int \Big(\sum_{k=1}^{K} u_k p_{kK}(w) \Big)^2 \mathbb{1}(0 < f_W(w) < t) \mathrm{d}w \leq \int \Big(\sum_{k=1}^{K} p_{kK}^2(w) \Big) \mathbb{1}(0 < f_W(w) < t) \mathrm{d}w$$

$$\leq \zeta_{0,K}^2 \int \mathbb{1}(0 < f_W(w) < t) \mathrm{d}w \leq C \zeta_{0,K}^2 t^\rho.$$

Take $t = c' \zeta_{0,K}^{-2/\rho}$ for some sufficiently small $c' > 0$ such that $C \zeta_{0,K}^2 t^\rho < 1/2$. We then obtain $\lambda_{\min}(\mathrm{E}[p_K(W_1) p_K(W_1)^\top]) \geq c' \zeta_{0,K}^{-2/\rho}/2$, as desired.

A.4 Proof of Proposition 4.1

For Part (i), W follows the distribution of the Gaussian copula from the multivariate normal distribution with correlation matrix Σ, and thus the Lebesgue density of W is

$$f_W(w) = \frac{1}{\sqrt{\det \Sigma}} \exp\Big(-\frac{1}{2} (\Phi^{-1}(w_1), \ldots, \Phi^{-1}(w_d))(\Sigma^{-1} - \boldsymbol{I}_d)(\Phi^{-1}(w_1), \ldots, \Phi^{-1}(w_d))^\top \Big),$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of a standard normal. Then,

$$\Big\{ w : 0 < f_W(w) < t \Big\}$$

$$= \Big\{ w : (\Phi^{-1}(w_1), \ldots, \Phi^{-1}(w_d))(\Sigma^{-1} - \boldsymbol{I}_d)(\Phi^{-1}(w_1), \ldots, \Phi^{-1}(w_d))^\top > 2 \log\Big(\frac{1}{t\sqrt{\det \Sigma}} \Big) \Big\}$$

$$\subset \Big\{ w : \|(\Phi^{-1}(w_1), \ldots, \Phi^{-1}(w_d))\|^2 > \frac{2}{\lambda_{\max}(\Sigma^{-1} - \boldsymbol{I}_d)} \log\Big(\frac{1}{t\sqrt{\det \Sigma}} \Big) \Big\}$$

$$\subset \bigcup_{k=1}^{d} \Big\{ w : \Phi^{-1}(w_k)^2 > \frac{2}{d\lambda_{\max}(\Sigma^{-1} - \boldsymbol{I}_d)} \log\Big(\frac{1}{t\sqrt{\det \Sigma}} \Big) \Big\}$$

$$= \bigcup_{k=1}^{d} \Big\{ w : w_k > \Phi\Big(\sqrt{\frac{2}{d\lambda_{\max}(\Sigma^{-1} - \boldsymbol{I}_d)} \log\Big(\frac{1}{t\sqrt{\det \Sigma}} \Big)} \Big) \Big\}.$$

For any $k \in [\![d]\!]$, by the Chernoff bound,

$$\mathsf{Leb}\Big(\Big\{ w : w_k > \Phi\Big(\sqrt{\frac{2}{d\lambda_{\max}(\Sigma^{-1} - \boldsymbol{I}_d)} \log\Big(\frac{1}{t\sqrt{\det \Sigma}} \Big)} \Big) \Big\} \Big)$$

$$= 1 - \Phi\Big(\sqrt{\frac{2}{d\lambda_{\max}(\Sigma^{-1} - \boldsymbol{I}_d)} \log\Big(\frac{1}{t\sqrt{\det \Sigma}} \Big)} \Big)$$

$$\leq \exp\Big(-\frac{1}{d\lambda_{\max}(\Sigma^{-1} - \boldsymbol{I}_d)} \log\Big(\frac{1}{t\sqrt{\det \Sigma}} \Big) \Big)$$

$$= \Big(t\sqrt{\det \Sigma} \Big)^{\frac{1}{d\lambda_{\max}(\Sigma^{-1} - \boldsymbol{I}_d)}}.$$

Then, we have

$$\mathsf{Leb}(\{w : 0 < f_W(w) < t\}) \le d\Big(t\sqrt{\det \Sigma}\Big)^{\frac{1}{d\lambda_{\max}(\Sigma^{-1}-I_d)}},$$

as desired.

A.5 Proof of Lemma 5.1

By simple algebra, we have

$$
\begin{aligned}
\widehat{\tau}_\phi &= \frac{1}{n}\sum_{i=1}^{n}\Big[\widehat{Y}_{\phi,i}(1) - \widehat{Y}_{\phi,i}(0)\Big] \\
&= \frac{1}{n}\sum_{i=1}^{n} D_i\Big[Y_i - \frac{1}{M}\sum_{j\in\mathcal{J}_\phi(i)}(Y_j + \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i}) - \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,j}))\Big] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}(1-D_i)\Big[\frac{1}{M}\sum_{j\in\mathcal{J}_\phi(i)}(Y_j + \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,j})) - Y_i\Big] \\
&= \frac{1}{n}\sum_{i=1,D_i=1}^{n}\Big[\widehat{R}_{\phi,i} + \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i}) - \frac{1}{M}\sum_{j\in\mathcal{J}_\phi(i)}\widehat{R}_{\phi,j}\Big] \\
&\quad + \frac{1}{n}\sum_{i=1,D_i=0}^{n}\Big[\frac{1}{M}\sum_{j\in\mathcal{J}_\phi(i)}\widehat{R}_{\phi,j} - \widehat{R}_{\phi,i} + \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i})\Big] \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big[\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i})\Big] + \frac{1}{n}\Big[\sum_{i=1,D_i=1}^{n}\Big(1+\frac{K_\phi(i)}{M}\Big)\widehat{R}_{\phi,i} - \sum_{i=1,D_i=0}^{n}\Big(1+\frac{K_\phi(i)}{M}\Big)\widehat{R}_{\phi,i}\Big].
\end{aligned}
$$

This completes the proof.

A.6 Proof of Theorem 5.1(i)

Part I. Suppose the propensity score model is correct, i.e., Assumption 5.2 and 5.3 hold. For any $i \in [\![n]\!]$, let $\bar{R}_{\phi,i} := Y_i - \bar{\mu}_{\phi,D_i}(U_{\phi,D_i,i})$. By Lemma 5.1,

$$
\begin{aligned}
\widehat{\tau}_\phi &= \widehat{\tau}_\phi^{\mathrm{reg}} + \frac{1}{n}\sum_{i=1}^{n}(2D_i - 1)\Big(1+\frac{K_\phi(i)}{M}\Big)\widehat{R}_{\phi,i} \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big[\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \bar{\mu}_{\phi,1}(U_{\phi,1,i})\Big] - \frac{1}{n}\sum_{i=1}^{n}\Big[\widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i}) - \bar{\mu}_{\phi,0}(U_{\phi,0,i})\Big] \\
&\quad + \frac{1}{n}\Big[\sum_{i=1}^{n}D_i\Big(1+\frac{K_\phi(i)}{M}\Big)\Big(\bar{\mu}_{\phi,1}(U_{\phi,1,i}) - \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i})\Big) - \sum_{i=1}^{n}(1-D_i)\Big(1+\frac{K_\phi(i)}{M}\Big)\Big(\bar{\mu}_{\phi,0}(U_{\phi,0,i}) - \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i})\Big)\Big] \\
&\quad + \frac{1}{n}\Big[\sum_{i=1}^{n}D_i\Big(1+\frac{K_\phi(i)}{M} - \frac{1}{e(X_i)}\Big)\bar{R}_{\phi,i} - \sum_{i=1}^{n}(1-D_i)\Big(1+\frac{K_\phi(i)}{M} - \frac{1}{1-e(X_i)}\Big)\bar{R}_{\phi,i}\Big] \\
&\quad + \frac{1}{n}\Big[\sum_{i=1}^{n}\Big(1-\frac{D_i}{e(X_i)}\Big)\bar{\mu}_{\phi,1}(U_{\phi,1,i}) - \sum_{i=1}^{n}\Big(1-\frac{1-D_i}{1-e(X_i)}\Big)\bar{\mu}_{\phi,0}(U_{\phi,0,i})\Big]
\end{aligned}
$$

21

$$+ \frac{1}{n} \Big[\sum_{i=1}^{n} \frac{D_i}{e(X_i)} Y_i - \sum_{i=1}^{n} \frac{1-D_i}{1-e(X_i)} Y_i \Big]. \tag{A.1}$$

For each pair of terms, we only establish the first half part under treatment, and the second half under control can be established in the same way.

For the first term in (A.1), by Assumption 5.2,

$$\Big| \frac{1}{n} \sum_{i=1}^{n} \Big[\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \bar{\mu}_{\phi,1}(U_{\phi,1,i}) \Big] \Big| \leq \Big| \frac{1}{n} \sum_{i=1}^{n} \Big[\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \bar{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) \Big] \Big| + \Big| \frac{1}{n} \sum_{i=1}^{n} \Big[\bar{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \bar{\mu}_{\phi,1}(U_{\phi,1,i}) \Big] \Big|$$

$$\leq \|\widehat{\mu}_1 - \bar{\mu}_1\|_\infty + \max_{i \in [\![n]\!]} \Big| \bar{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \bar{\mu}_{\phi,1}(U_{\phi,1,i}) \Big| = o_\mathrm{P}(1).$$

Then

$$\frac{1}{n} \sum_{i=1}^{n} \Big[\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \bar{\mu}_{\phi,1}(U_{\phi,1,i}) \Big] = o_\mathrm{P}(1). \tag{A.2}$$

For the second term in (A.1), by Assumption 5.2,

$$\Big| \frac{1}{n} \sum_{i=1}^{n} D_i \Big(1 + \frac{K_\phi(i)}{M} \Big) \Big(\bar{\mu}_{\phi,1}(U_{\phi,1,i}) - \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) \Big) \Big|$$

$$\leq \max_{i \in [\![n]\!]} \Big| \bar{\mu}_{\phi,1}(U_{\phi,1,i}) - \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) \Big| \cdot \frac{1}{n} \sum_{i=1}^{n} D_i \Big(1 + \frac{K_\phi(i)}{M} \Big) = \max_{i \in [\![n]\!]} \Big| \bar{\mu}_{\phi,1}(U_{\phi,1,i}) - \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) \Big|$$

$$\leq \|\widehat{\mu}_1 - \bar{\mu}_1\|_\infty + \max_{i \in [\![n]\!]} \Big| \bar{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \bar{\mu}_{\phi,1}(U_{\phi,1,i}) \Big| = o_\mathrm{P}(1).$$

We then have

$$\frac{1}{n} \sum_{i=1}^{n} D_i \Big(1 + \frac{K_\phi(i)}{M} \Big) \Big(\bar{\mu}_{\phi,1}(U_{\phi,1,i}) - \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) \Big) = o_\mathrm{P}(1). \tag{A.3}$$

For the third term in (A.1), by the Cauchy-Schwarz inequality,

$$\Big| \frac{1}{n} \sum_{i=1}^{n} D_i \Big(1 + \frac{K_\phi(i)}{M} - \frac{1}{e(X_i)} \Big) \bar{R}_{\phi,i} \Big|$$

$$\leq \Big\{ \frac{1}{n} \sum_{i=1}^{n} D_i \Big(1 + \frac{K_\phi(i)}{M} - \frac{1}{e(X_i)} \Big)^2 \Big\}^{1/2} \Big\{ \frac{1}{n} \sum_{i=1}^{n} D_i \bar{R}_{\phi,i}^2 \Big\}^{1/2}.$$

Note that by Assumptions 5.1 and 5.2,

$$\mathrm{E}\Big[\frac{1}{n} \sum_{i=1}^{n} D_i \bar{R}_{\phi,i}^2 \Big] = \mathrm{E}\Big[D_1 \bar{R}_{\phi,1}^2 \Big] = \mathrm{E}\Big[D_1 \Big(Y_1(1) - \bar{\mu}_{\phi,1}(U_{\phi,1,1}) \Big)^2 \Big]$$

$$\leq 2\mathrm{E}\Big[D_1 \Big(\sigma_1^2(U_{\phi,1,1}) + [\mu_{\phi,1}(U_{\phi,1,1}) - \bar{\mu}_{\phi,1}(U_{\phi,1,1})]^2 \Big) \Big] < \infty,$$

where $\sigma_1^2(u) := \mathrm{E}[(Y(1) - \mu_{\phi,1}(u))^2 \,|\, U_{\phi,1} = u]$ for $u \in \mathcal{X}_\phi$. We then obtain by Assumption 5.3 and the Markov inequality that

$$\frac{1}{n} \sum_{i=1}^{n} D_i \Big(1 + \frac{K_\phi(i)}{M} - \frac{1}{e(X_i)} \Big) \bar{R}_{\phi,i} = o_\mathrm{P}(1). \tag{A.4}$$

22

For the fourth term in (A.1), notice that $\bar{\mu}_{\phi,1}(U_{\phi,1,i})$ is a function of X_i. Then by the definition of the propensity score and Assumption 5.1,

$$\mathrm{E}\Big[\Big(1 - \frac{D_i}{e(X_i)}\Big)\bar{\mu}_{\phi,1}(U_{\phi,1,i})\Big] = 0, \quad \mathrm{E}\Big[\Big|\Big(1 - \frac{D_i}{e(X_i)}\Big)\bar{\mu}_{\phi,1}(U_{\phi,1,i})\Big|\Big] < \infty.$$

By the i.i.d of $[(X_i, D_i)]_{i=1}^{n}$ and the weak law of large numbers, we have

$$\frac{1}{n}\sum_{i=1}^{n}\Big(1 - \frac{D_i}{e(X_i)}\Big)\bar{\mu}_{\phi,1}(U_{\phi,1,i}) = o_{\mathrm{P}}(1). \tag{A.5}$$

For the fifth term in (A.1), notice that $\mathrm{E}[\|Y\|]$ is bounded from Assumption 3.1 and $[(X_i, D_i, Y_i)]_{i=1}^{n}$ are i.i.d.. Using the weak law of large numbers yields

$$\frac{1}{n}\Big[\sum_{i=1}^{n}\frac{D_i}{e(X_i)}Y_i - \sum_{i=1}^{n}\frac{1 - D_i}{1 - e(X_i)}Y_i\Big] \xrightarrow{\mathrm{p}} \mathrm{E}\Big[Y_i(1) - Y_i(0)\Big] = \tau. \tag{A.6}$$

Plugging (A.2), (A.3), (A.4), (A.5) into (A.1) completes the proof.

Part II. Suppose the outcome model is correct, i.e., Assumption 5.4 holds. Using the representation (3.1),

$$\widehat{\tau}_{\phi} = \widehat{\tau}_{\phi}^{\mathrm{reg}} + \frac{1}{n}\sum_{i=1}^{n}(2D_i - 1)\Big(1 + \frac{K_{\phi}(i)}{M}\Big)\widehat{R}_{\phi,i}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\Big[\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \mu_{\phi,1}(U_{\phi,1,i})\Big] - \frac{1}{n}\sum_{i=1}^{n}\Big[\widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i}) - \mu_{\phi,0}(U_{\phi,0,i})\Big]$$

$$+ \frac{1}{n}\Big[\sum_{i=1}^{n}D_i\Big(1 + \frac{K_{\phi}(i)}{M}\Big)\Big(\mu_{\phi,1}(U_{\phi,1,i}) - \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i})\Big) - \sum_{i=1}^{n}(1 - D_i)\Big(1 + \frac{K_{\phi}(i)}{M}\Big)\Big(\mu_{\phi,0}(U_{\phi,0,i}) - \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i})\Big)\Big]$$

$$+ \frac{1}{n}\Big[\sum_{i=1}^{n}D_i\Big(1 + \frac{K_{\phi}(i)}{M}\Big)\Big(Y_i - \mu_{\phi,1}(U_{\phi,1,i})\Big) - \sum_{i=1}^{n}(1 - D_i)\Big(1 + \frac{K_{\phi}(i)}{M}\Big)\Big(Y_i - \mu_{\phi,0}(U_{\phi,0,i})\Big)\Big]$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\Big[\mu_{\phi,1}(U_{\phi,1,i}) - \mu_{\phi,0}(U_{\phi,0,i})\Big]. \tag{A.7}$$

For the first term in (A.7), in the same way as (A.2),

$$\frac{1}{n}\sum_{i=1}^{n}\Big[\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \mu_{\phi,1}(U_{\phi,1,i})\Big] = o_{\mathrm{P}}(1). \tag{A.8}$$

For the second term in (A.7), in the same way as (A.3),

$$\frac{1}{n}\sum_{i=1}^{n}D_i\Big(1 + \frac{K_{\phi}(i)}{M}\Big)\Big(\mu_{\phi,1}(U_{\phi,1,i}) - \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i})\Big) = o_{\mathrm{P}}(1). \tag{A.9}$$

For the third term in (A.7), noticing that $[K_{\phi}(i)]_{i=1}^{n}$ is a function of $\{(X_i, D_i)\}_{i=1}^{n}$, by Assumption 5.1 and Assumption 5.4, we can obtain for any $i \in [\![n]\!]$,

$$\mathrm{E}\Big[D_i\Big(1 + \frac{K_{\phi}(i)}{M}\Big)\Big(Y_i - \mu_{\phi,1}(U_{\phi,1,i})\Big)\,\Big|\,\{(X_i, D_i)\}_{i=1}^{n}\Big]$$

$$= D_i\Big(1 + \frac{K_{\phi}(i)}{M}\Big)\Big(\mathrm{E}[Y_i \,|\, X_i, D_i = 1] - \mu_{\phi,1}(U_{\phi,1,i})\Big)$$

$$= D_i \Big(1 + \frac{K_\phi(i)}{M} \Big) \Big(\mathrm{E}[Y_i(1) \,|\, X_i] - \mu_{\phi,1}(U_{\phi,1,i}) \Big) = 0,$$

and

$$\mathrm{E}\Big[\Big| \frac{1}{n} \sum_{i=1}^n D_i \Big(1 + \frac{K_\phi(i)}{M} \Big) \Big(Y_i - \mu_{\phi,1}(U_{\phi,1,i}) \Big) \Big| \Big] \leq \mathrm{E}\Big[\Big| \frac{1}{n} \sum_{i=1}^n D_i \Big(1 + \frac{K_\phi(i)}{M} \Big) \Big| \Big] \|\sigma_1\|_\infty$$

$$\lesssim \|\sigma_1\|_\infty = O(1).$$

Accordingly, by the martingale convergence theorem in the same way as Abadie and Imbens (2012), we obtain

$$\frac{1}{n} \sum_{i=1}^n D_i \Big(1 + \frac{K_\phi(i)}{M} \Big) \Big(Y_i - \mu_{\phi,1}(U_{\phi,1,i}) \Big) = o_{\mathrm{P}}(1). \tag{A.10}$$

For the fourth term in (A.7), notice that $\mathrm{E}[\mu_{\phi,\omega}^2(U_{\phi,\omega})]$ is bounded for $\omega = 0, 1$. Using the weak law of large number, we obtain

$$\frac{1}{n} \sum_{i=1}^n \Big[\mu_{\phi,1}(U_{\phi,1,i}) - \mu_{\phi,0}(U_{\phi,0,i}) \Big] \xrightarrow{\mathrm{p}} \mathrm{E}\Big[\mu_1(X_1) - \mu_0(X_1) \Big] = \tau. \tag{A.11}$$

Plugging (A.8), (A.9), (A.10), (A.11) into (A.7) completes the proof.

A.7   Proof of Theorem 5.1(ii)

We decompose $\widehat{\tau}_\phi$ as

$$\begin{aligned}
\widehat{\tau}_\phi &= \widehat{\tau}_\phi^{\mathrm{reg}} + \frac{1}{n} \sum_{i=1}^n (2D_i - 1) \Big(1 + \frac{K_\phi(i)}{M} \Big) \widehat{R}_{\phi,i} \\
&= \frac{1}{n} \sum_{i=1}^n \Big[\mu_{\phi,1}(U_{\phi,1,i}) - \mu_{\phi,0}(U_{\phi,0,i}) \Big] + \frac{1}{n} \sum_{i=1}^n (2D_i - 1) \Big(1 + \frac{K_\phi(i)}{M} \Big) \Big(Y_i - \mu_{\phi,D_i}(U_{\phi,D_i,i}) \Big) \\
&\quad + \frac{1}{n} \sum_{i=1}^n (2D_i - 1) \Big[\mu_{\phi,1-D_i}(U_{\phi,1-D_i,i}) - \frac{1}{M} \sum_{j \in \mathcal{J}_\phi(i)} \mu_{\phi,1-D_i}(U_{\phi,1-D_i,j}) \Big] \\
&\quad - \frac{1}{n} \sum_{i=1}^n (2D_i - 1) \Big[\widehat{\mu}_{\phi,1-D_i}(\widehat{U}_{\phi,1-D_i,i}) - \frac{1}{M} \sum_{j \in \mathcal{J}_\phi(i)} \widehat{\mu}_{\phi,1-D_i}(\widehat{U}_{\phi,1-D_i,j}) \Big] \\
&=: \bar{\tau}_\phi + E_n + B_n - \widehat{B}_n.
\end{aligned}$$

In the same way as Lemmas A.1 and A.2 in Lin and Han (2022), we have the following central limit theorem on $\bar{\tau}_\phi + E_n$.

Lemma A.1. *Under Assumptions 5.1, 5.3, 5.5,*

$$\sqrt{n} \sigma^{-1} \Big(\bar{\tau}_\phi + E_n - \tau \Big) \xrightarrow{\mathrm{d}} N \Big(0, 1 \Big).$$

For the bias term $B_M - \widehat{B}_M$, in light of the smoothness conditions on μ_ω and approximation conditions on $\widehat{\mu}_\omega$ for $\omega = 0, 1$, one can establish the following lemma.

Lemma A.2. *Under Assumptions 5.1, 5.3, 5.5, 5.6, 5.7,*
$$\sqrt{n}\Big(B_n - \widehat{B}_n\Big) \xrightarrow{\ \mathsf{p}\ } 0.$$

Combining Lemma A.1 and Lemma A.2 completes the proof.

The consistency of the variance estimator can be established in a similar way as the proof of Theorem 4.1 in Lin et al. (2023).

A.8   Proof of Theorem 5.2

Similar to Lin et al. (2023), we first consider a two-sample density ratio estimation problem.

With an abuse of notation, restricted to this section let's consider two general random vectors X, Z in $\mathbb{R}^d$ that are defined on the same probability space. Let $\mathcal{X}, \mathcal{Z} \subset \mathbb{R}^d$ be the supports of X and Z respectively with $\mathcal{Z} \subset \mathcal{X}$. Consider a general function $\phi : \mathcal{X} \to \mathbb{R}^m$. Let ν_0 and ν_1 represent the probability measures of $\phi(X)$ and $\phi(Z)$, respectively. Assume ν_0 and ν_1 are absolutely continuous with respect to the Lebesgue measure λ on $\mathbb{R}^m$ equipped with the Euclidean norm $\|\cdot\|$; denote the corresponding densities (Radon-Nikodym derivatives) by f_0 and f_1. Assume further that ν_1 is absolutely continuous with respect to ν_0 and write the corresponding density ratio, f_1/f_0, as r; set $0/0 = 0$ by default.

Assume $X_1, \ldots, X_{N_0}$ are N_0 independent copies of X, $Z_1, \ldots, Z_{N_1}$ are N_1 independent copies of Z, and $[X_i]_{i=1}^{N_0}$ and $[Z_j]_{j=1}^{N_1}$ are mutually independent. We aim to estimate the density ratio $r(\phi(x))$ for any $x \in \mathcal{X}$ based on $\{X_1, \ldots, X_{N_0}, Z_1, \ldots, Z_{N_1}\}$.

For any $x \in \mathcal{X}$, we consider a general estimator $\widehat{\phi}$ estimating ϕ, which may depend on

$$\{X_1, \ldots, X_{N_0}, Z_1, \ldots, Z_{N_1}\} \text{ and } x.$$

Define the catchment area of x:

$$A_\phi(x) = A_\phi\big(x, \{X_i\}_{i=1}^{N_0}, \widehat{\phi}\big) := \Big\{ z \in \mathcal{Z} : \|\widehat{\phi}(x) - \widehat{\phi}(z)\| \le \widehat{\Phi}_M(z) \Big\}, \tag{A.12}$$

where $\widehat{\Phi}_M(z)$ is the M-th order statistics of $\{\|\widehat{\phi}(X_i) - \widehat{\phi}(z)\|\}_{i=1}^{N_0}$, and the number of matched times of x:

$$K_\phi(x) = K_\phi\big(x, \{X_i\}_{i=1}^{N_0}, \{Z_j\}_{j=1}^{N_1}\big) := \sum_{j=1}^{N_1} \mathbb{1}\Big(Z_j \in A_\phi(x)\Big). \tag{A.13}$$

Then the density ratio estimator is defined as:

$$\widehat{r}_\phi(x) = \widehat{r}_\phi\big(x, \{X_i\}_{i=1}^{N_0}, \{Z_j\}_{j=1}^{N_1}\big) := \frac{N_0}{N_1} \frac{K_\phi(x)}{M}. \tag{A.14}$$

For any positive integer p, let $\widehat{\phi}_{(Z_1,\ldots,Z_p)\to z}$ be the estimator replacing $(Z_1, \ldots, Z_p)$ by z for $z \in \mathcal{Z}^p$. We consider the following two assumptions, which are analogies of Assumption 5.9 and Assumption 5.10 in the two-sample problem.

Assumption A.1.
$$\lim_{N_0 \to \infty} \mathrm{E}\Big[\Big(\frac{N_0}{M}\Big)^p \sup_{z \in \mathcal{Z}^p} \|\widehat{\phi}_{(Z_1,\ldots,Z_p)\to z} - \phi\|_\infty^{pd}\Big] = 0.$$

Assumption A.2. *For any $\epsilon > 0$ and $\delta > 0$,*

$$\mathrm{P}\Big(\sup_{\delta \in \mathbb{R}: \delta \geq u} \delta^{-1} \sup_{s,t \in \mathcal{X}: \|\phi(s)-\phi(t)\| \leq \delta} \sup_{z \in \mathcal{Z}^p} \|(\widehat{\phi}_{(Z_1,\ldots,Z_p) \to z} - \phi)(s) - (\widehat{\phi}_{(Z_1,\ldots,Z_p) \to z} - \phi)(t)\| > \epsilon \Big) \leq T_\epsilon(u),$$

for $T_\epsilon(u)$ satisfying for any $k \in [\![p]\!]$,

$$\lim_{N_0 \to \infty} \Big(\frac{N_0}{M} \Big)^k \int_0^\infty u^{k-1} T_\epsilon(u^{1/m}) \mathrm{d}u = 0,$$

and

$$\lim_{N_0 \to \infty} \Big(\frac{N_0}{M} \Big)^p \mathrm{P}\Big(\|\widehat{\phi} - \phi\|_\infty > \epsilon \Big) = 0.$$

The following theorem considers the asymptotic L^p moments of $\widehat{r}_\phi$.

Theorem A.1 (Asymptotic L^p moments of $\widehat{r}_\phi$). *Let p be any positive integer. Assume Assumption A.1 or A.2 holds for p. Assume $M \log N_0 / N_0 \to 0$, $M N_1 / N_0 \to \infty$ and $M \to \infty$ as $N_0 \to \infty$. We then have*

$$\lim_{N_0 \to \infty} \mathrm{E}\big[(\widehat{r}_\phi(x))^p \big] = \big[r(\phi(x)) \big]^p$$

holds for all $x \in \mathcal{X}$ such that $f_0(\phi(x)) > 0$ and f_0, f_1 are continuous at $\phi(x)$.

The proof of Theorem A.1 will use the following lemma.

Lemma A.3. *Under the same conditions of Theorem A.1, we have*

$$\lim_{N_0 \to \infty} \Big(\frac{N_0}{M} \Big)^p \mathrm{P}\Big(Z_1, \ldots, Z_p \in A_\phi(x) \Big) = \big[r(\phi(x)) \big]^p.$$

holds for all $x \in \mathcal{X}$ such that $f_0(\phi(x)) > 0$ and f_0, f_1 are continuous at $\phi(x)$.

As a direct result of Theorem A.1, we can establish the pointwise consistency of the estimator $\widehat{r}_\phi$.

Corollary A.1 (Pointwise consistency). *Under the same conditions as Theorem A.1, if p is even, we have*

$$\lim_{N_0 \to \infty} \mathrm{E}\big[|\widehat{r}_\phi(x) - r(\phi(x))|^p \big] = 0$$

holds for all $x \in \mathcal{X}$ such that $f_0(\phi(x)) > 0$ and f_0, f_1 are continuous at $\phi(x)$.

The pointwise consistency of $\widehat{r}_\phi$ can then be generalized to global consistency under the following assumptions on $\mathcal{X}$.

Assumption A.3. *(i) $\mathcal{X}$ is compact and the surface areas of $\mathcal{X}$ and $\mathcal{Z}$ are bounded.*

(ii) r is bounded over $\mathcal{X}$.

(iii) f_0 is continuous over $\mathcal{X}$ and f_1 is continuous over $\mathcal{Z}$.

Theorem A.2 (Global consistency). *Under the same conditions of Theorem A.1 and Assumption A.3, if p is even, we have*

$$\lim_{N_0 \to \infty} \mathrm{E}\Big[\Big| \widehat{r}_\phi(X) - r(\phi(X)) \Big|^p \Big] = 0.$$

26

Now back to the causal setting, the density ratio $r(\phi(x))$ is not necessarily equal to the density ratio we are interested in. We consider a lemma showing the equivalence of the two density ratios under additional assumptions.

Lemma A.4. *Let $f_{X\,|\,D=1}$ and $f_{X\,|\,D=0}$ be the density of $X\,|\,D=1$ and $X\,|\,D=0$, respectively. Let $f_{\phi,X\,|\,D=1}$ and $f_{\phi,X\,|\,D=0}$ be the density of $\phi(X)\,|\,D=1$ and $\phi(X)\,|\,D=0$. Then for any $x \in \mathcal{X}$ such that $\mathrm{P}(D=1\,|\,\phi(X)=\phi(x)) = \mathrm{P}(D=1\,|\,X=x)$, we have*

$$\frac{f_{\phi,X\,|\,D=1}(\phi(x))}{f_{\phi,X\,|\,D=0}(\phi(x))} = \frac{f_{X\,|\,D=1}(x)}{f_{X\,|\,D=0}(x)}.$$

Note that

$$\mathrm{E}\Big[\frac{K_\phi(1)}{M} - \Big(D_1\frac{1-e(X_1)}{e(X_1)} + (1-D_1)\frac{e(X_1)}{1-e(X_1)}\Big)\Big]^2$$

$$= \mathrm{E}\Big[\mathrm{E}\Big[\Big[\frac{K_\phi(1)}{M} - \Big(D_1\frac{1-e(X_1)}{e(X_1)} + (1-D_1)\frac{e(X_1)}{1-e(X_1)}\Big)\Big]^2 \,\Big|\, \{D_i\}_{i=1}^n\Big]\Big]$$

$$= \mathrm{E}\Big[\mathrm{E}\Big[\Big(\frac{K_\phi(1)}{M} - \frac{1-e(X_1)}{e(X_1)}\Big)^2 \,\Big|\, \{D_i\}_{i=1}^n, D_1 = 1\Big]\mathbb{1}\Big(D_1 = 1\Big)\Big]$$

$$+ \mathrm{E}\Big[\mathrm{E}\Big[\Big(\frac{K_\phi(1)}{M} - \frac{e(X_1)}{1-e(X_1)}\Big)^2 \,\Big|\, \{D_i\}_{i=1}^n, D_1 = 0\Big]\mathbb{1}\Big(D_1 = 0\Big)\Big].$$

We consider the second term for example. Conditional on $\{D_i\}_{i=1}^n$, $[X_i]_{i:D_i=0}$ and $[X_i]_{i:D_i=1}$ are two samples from $X\,|\,D=0$ and $X\,|\,D=1$, respectively. Note that

$$\mathrm{E}\Big[\Big(\frac{K_\phi(1)}{M} - \frac{e(X_1)}{1-e(X_1)}\Big)^2 \,\Big|\, \{D_i\}_{i=1}^n, D_1 = 0\Big] = \Big(\frac{N_1}{N_0}\Big)^2\mathrm{E}\Big[\Big(\frac{N_0}{N_1}\frac{K_\phi(1)}{M} - \frac{N_0}{N_1}\frac{e(X_1)}{1-e(X_1)}\Big)^2 \,\Big|\, \{D_i\}_{i=1}^n, D_1 = 0\Big].$$

By the strong law of large number, we have $(N_0/N_1) \to \mathrm{P}(D=0)/\mathrm{P}(D=1)$ with probability one. Note that

$$\frac{\mathrm{P}(D=0)}{\mathrm{P}(D=1)}\frac{e(X_1)}{1-e(X_1)} = \frac{f_{X\,|\,D=1}(x)}{f_{X\,|\,D=0}(x)}.$$

To apply Theorem A.2 and Lemma A.4, the last thing is to compare the definition of $K_\phi(1)$ with $K_\phi(x)$. Note that if we define

$$A'_\phi(x) := \Big\{z \in \mathcal{Z} : \|\widehat{\phi}(x) - \widehat{\phi}(z)\| < \widehat{\Phi}_M(z)\Big\}, \quad K'_\phi(x) := \sum_{j=1}^{N_1}\mathbb{1}\Big(Z_j \in A'_\phi(x)\Big),$$

as long as the ties are broken in arbitrary way, we can check that $K'_\phi(X_1) \le K_\phi(1) \le K_\phi(X_1)$. Note that all the previous results for $K_\phi(x)$ are also hold for $K'_\phi(x)$. Then the proof is complete.

B   Proofs of Auxiliary Results

B.1   Proof of Lemma A.2

Note that for any $i \in [\![n]\!]$ and $\omega = 0, 1$,

$$\Big|\mu_{\phi,\omega}(U_{\phi,\omega,i}) - \mu_{\phi,\omega}(U_{\phi,\omega,j}) - \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) + \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,j})\Big|$$

$$\leq \left| \mu_{\phi,\omega}(U_{\phi,\omega,i}) - \mu_{\phi,\omega}(U_{\phi,\omega,j}) - \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,i}) + \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,j}) \right|$$
$$+ \left| \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,i}) - \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,j}) - \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) + \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,j}) \right|. \tag{B.1}$$

We can also decompose in another way:
$$\left| \mu_{\phi,\omega}(U_{\phi,\omega,i}) - \mu_{\phi,\omega}(U_{\phi,\omega,j}) - \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) + \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,j}) \right|$$
$$\leq \left| \mu_{\phi,\omega}(U_{\phi,\omega,i}) - \mu_{\phi,\omega}(U_{\phi,\omega,j}) - \mu_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) + \mu_{\phi,\omega}(\widehat{U}_{\phi,\omega,j}) \right|$$
$$+ \left| \mu_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) - \mu_{\phi,\omega}(\widehat{U}_{\phi,\omega,j}) - \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) + \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,j}) \right|.$$

We consider the proof under (B.1), and the proof under the second decomposition is similar.

For the first term in (B.1), by Taylor expansion to k-th order with $k = \max\{\lfloor m/2 \rfloor, 1\} + 1$,
$$\left| \mu_{\phi,\omega}(U_{\phi,\omega,j}) - \mu_{\phi,\omega}(U_{\phi,\omega,i}) - \sum_{\ell=1}^{k-1} \sum_{t \in \Lambda_\ell} \frac{1}{t!} \partial^t \mu_{\phi,\omega}(U_{\phi,\omega,i})(U_{\phi,\omega,j} - U_{\phi,\omega,i})^t \right|$$
$$\leq \max_{t \in \Lambda_k} \|\partial^t \mu_{\phi,\omega}\|_\infty \sum_{t \in \Lambda_k} \frac{1}{t!} \|U_{\phi,\omega,j} - U_{\phi,\omega,i}\|^k.$$

In the same way,
$$\left| \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,j}) - \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,i}) - \sum_{\ell=1}^{k-1} \sum_{t \in \Lambda_\ell} \frac{1}{t!} \partial^t \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,i})(U_{\phi,\omega,j} - U_{\phi,\omega,i})^t \right|$$
$$\leq \max_{t \in \Lambda_k} \|\partial^t \widehat{\mu}_{\phi,\omega}\|_\infty \sum_{t \in \Lambda_k} \frac{1}{t!} \|U_{\phi,\omega,j} - U_{\phi,\omega,i}\|^k.$$

We also have
$$\left| \sum_{\ell=1}^{k-1} \sum_{t \in \Lambda_\ell} \frac{1}{t!} (\partial^t \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,i}) - \partial^t \mu_{\phi,\omega}(U_{\phi,\omega,i}))(U_{\phi,\omega,j} - U_{\phi,\omega,i})^t \right|$$
$$\leq \sum_{\ell=1}^{k-1} \max_{t \in \Lambda_\ell} \|\partial^t \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,i}) - \partial^t \mu_{\phi,\omega}(U_{\phi,\omega,i})\| \sum_{t \in \Lambda_\ell} \frac{1}{t!} \|U_{\phi,\omega,j} - U_{\phi,\omega,i}\|^\ell.$$

For the second term in (B.1), by Taylor expansion,
$$\left| \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,i}) - \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,j}) - \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) + \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,j}) \right|$$
$$= \left| \partial \widehat{\mu}_{\phi,\omega}(\bar{u}_j)^\top (\widehat{U}_{\phi,\omega,j} - U_{\phi,\omega,j}) - \partial \widehat{\mu}_{\phi,\omega}(\bar{u}_i)^\top (\widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,i}) \right|$$
$$\leq \left| [\partial \widehat{\mu}_{\phi,\omega}(\bar{u}_j) - \partial \widehat{\mu}_{\phi,\omega}(\bar{u}_i)]^\top (\widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,i}) \right| + \left| \partial \widehat{\mu}_{\phi,\omega}(\bar{u}_j)^\top (\widehat{U}_{\phi,\omega,j} - \widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,j} + U_{\phi,\omega,i}) \right|$$
$$\lesssim \max_{t \in \Lambda_2} \|\partial^t \widehat{\mu}_{\phi,\omega}\|_\infty \|\bar{u}_j - \bar{u}_i\| \|\widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,i}\| + \|\partial \widehat{\mu}_{\phi,\omega}\|_\infty \|\widehat{U}_{\phi,\omega,j} - \widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,j} + U_{\phi,\omega,i}\|,$$

where $\bar{u}_i$ is between $U_{\phi,\omega,i}$ and $\widehat{U}_{\phi,\omega,i}$, and $\bar{u}_j$ is between $U_{\phi,\omega,j}$ and $\widehat{U}_{\phi,\omega,j}$. Since $\|\widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,i}\| \leq \|\widehat{\phi}_\omega - \phi_\omega\|_\infty$ and $\|\bar{u}_j - \bar{u}_i\| \leq \|U_{\phi,\omega,j} - U_{\phi,\omega,i}\| + \|\widehat{\phi}_\omega - \phi_\omega\|_\infty$, we have
$$\left| \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,i}) - \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,j}) - \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) + \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,j}) \right|$$

$$\lesssim \max_{t\in\Lambda_2}\|\partial^t\widehat{\mu}_{\phi,\omega}\|_\infty[\|\widehat{\phi}_\omega - \phi_\omega\|_\infty^2 + \|\widehat{\phi}_\omega - \phi_\omega\|_\infty\|U_{\phi,\omega,j} - U_{\phi,\omega,i}\|] + \|\partial\widehat{\mu}_{\phi,\omega}\|_\infty\|\widehat{U}_{\phi,\omega,j} - \widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,j} + U_{\phi,\omega,i}\|.$$

Since we have $|\mathcal{J}_\phi(i)| = M$ for any $i \in [\![n]\!]$, then

$$|B_n - \widehat{B}_n|$$

$$\leq \frac{1}{n}\sum_{i=1}^n \frac{1}{M}\sum_{j\in\mathcal{J}_\phi(i)}\left|\mu_{\phi,1-D_i}(U_{\phi,1-D_i,i}) - \mu_{\phi,1-D_i}(U_{\phi,1-D_i,j}) - \widehat{\mu}_{\phi,1-D_i}(\widehat{U}_{\phi,1-D_i,i}) + \widehat{\mu}_{\phi,1-D_i}(\widehat{U}_{\phi,1-D_i,j})\right|$$

$$\leq \frac{1}{n}\sum_{i=1}^n \max_{j\in\mathcal{J}_\phi(i)}\left|\mu_{\phi,1-D_i}(U_{\phi,1-D_i,i}) - \mu_{\phi,1-D_i}(U_{\phi,1-D_i,j}) - \widehat{\mu}_{\phi,1-D_i}(\widehat{U}_{\phi,1-D_i,i}) + \widehat{\mu}_{\phi,1-D_i}(\widehat{U}_{\phi,1-D_i,j})\right|$$

$$\lesssim \max_{\omega\in\{0,1\}}\left(\max_{t\in\Lambda_k}\|\partial^t\mu_{\phi,\omega}\|_\infty + \max_{t\in\Lambda_k}\|\partial^t\widehat{\mu}_{\phi,\omega}\|_\infty\right)\left(\frac{1}{n}\sum_{i=1}^n\max_{j\in\mathcal{J}_\phi(i)}\|U_{\phi,1-D_i,j} - U_{\phi,1-D_i,i}\|^k\right)$$

$$+ \sum_{\ell=1}^{k-1}\left(\frac{1}{n}\sum_{i=1}^n\max_{t\in\Lambda_\ell}\|\partial^t\widehat{\mu}_{\phi,1-D_i}(U_{\phi,1-D_i,i}) - \partial^t\mu_{\phi,1-D_i}(U_{\phi,1-D_i,i})\|\max_{j\in\mathcal{J}_\phi(i)}\|U_{\phi,1-D_i,j} - U_{\phi,1-D_i,i}\|^\ell\right)$$

$$+ \max_{\omega\in\{0,1\}}\max_{t\in\Lambda_2}\|\partial^t\widehat{\mu}_{\phi,\omega}\|_\infty\left[\|\widehat{\phi}_\omega - \phi_\omega\|_\infty^2 + \|\widehat{\phi}_\omega - \phi_\omega\|_\infty\left(\frac{1}{n}\sum_{i=1}^n\max_{j\in\mathcal{J}_\phi(i)}\|U_{\phi,1-D_i,j} - U_{\phi,1-D_i,i}\|\right)\right]$$

$$+ \max_{\omega\in\{0,1\}}\|\partial\widehat{\mu}_{\phi,\omega}\|_\infty\left(\frac{1}{n}\sum_{i=1}^n\max_{j\in\mathcal{J}_\phi(i)}\|\widehat{U}_{\phi,1-D_i,j} - \widehat{U}_{\phi,1-D_i,i} - U_{\phi,1-D_i,j} + U_{\phi,1-D_i,i}\|\right). \tag{B.2}$$

For any $i \in [\![n]\!]$, let $\widetilde{\mathcal{J}}_\phi(i)$ be the index set of M-NNs of $U_{\phi,1-D_i,i}$ in $\{U_{\phi,1-D_i,j} : D_j = 1-D_i\}_{j=1}^n$ with ties broken in arbitrary way. Then

$$\max_{j\in\mathcal{J}_\phi(i)}\|U_{\phi,1-D_i,j} - U_{\phi,1-D_i,i}\| \leq \max_{j\in\mathcal{J}_\phi(i)}\|\widehat{U}_{\phi,1-D_i,j} - \widehat{U}_{\phi,1-D_i,i}\| + 2\|\widehat{\phi}_{1-D_i} - \phi_{1-D_i}\|_\infty$$

$$\leq \max_{j\in\widetilde{\mathcal{J}}_\phi(i)}\|\widehat{U}_{\phi,1-D_i,j} - \widehat{U}_{\phi,1-D_i,i}\| + 2\|\widehat{\phi}_{1-D_i} - \phi_{1-D_i}\|_\infty$$

$$\leq \max_{j\in\widetilde{\mathcal{J}}_\phi(i)}\|U_{\phi,1-D_i,j} - U_{\phi,1-D_i,i}\| + 4\|\widehat{\phi}_{1-D_i} - \phi_{1-D_i}\|_\infty.$$

By Li and Racine (2023, Lemma 14.1), as long as the density of $U_{\phi,\omega}$ is continuous for $\omega = 0, 1$, we have for any positive integer p,

$$\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^n\max_{j\in\widetilde{\mathcal{J}}_\phi(i)}\|U_{\phi,1-D_i,j} - U_{\phi,1-D_i,i}\|^p\right] \lesssim \left(\frac{M}{n}\right)^{p/m}.$$

Then for any positive integer p, by Assumption 5.7, we have

$$\frac{1}{n}\sum_{i=1}^n\max_{j\in\mathcal{J}_\phi(i)}\|U_{\phi,1-D_i,j} - U_{\phi,1-D_i,i}\|^p$$

$$\lesssim \frac{1}{n}\sum_{i=1}^n\left(\max_{j\in\widetilde{\mathcal{J}}_\phi(i)}\|U_{\phi,1-D_i,j} - U_{\phi,1-D_i,i}\|^p + \|\widehat{\phi}_{1-D_i} - \phi_{1-D_i}\|_\infty^p\right)$$

$$= O_\mathrm{P}((M/n)^{p/m} + n^{-p/2}). \tag{B.3}$$

For any positive integer $\ell \in [\![k-1]\!]$, by Assumption 5.6, we have

$$\frac{1}{n}\sum_{i=1}^{n}\max_{t\in\Lambda_\ell}\|\partial^t\widehat{\mu}_{\phi,1-D_i}(U_{\phi,1-D_i,i}) - \partial^t\mu_{\phi,1-D_i}(U_{\phi,1-D_i,i})\|\max_{j\in\mathcal{J}_\phi(i)}\|U_{\phi,1-D_i,j} - U_{\phi,1-D_i,i}\|^\ell$$

$$\leq \max_{\omega\in\{0,1\}}\max_{t\in\Lambda_\ell}\|\partial^t\widehat{\mu}_{\phi,\omega} - \partial^t\mu_{\phi,\omega}\|_\infty\Big(\frac{1}{n}\sum_{i=1}^{n}\max_{j\in\mathcal{J}_\phi(i)}\|U_{\phi,1-D_i,j} - U_{\phi,1-D_i,i}\|^\ell\Big)$$

$$= O_P(n^{-\gamma_\ell}((M/n)^{\ell/m} + n^{-\ell/2})). \tag{B.4}$$

For any $\epsilon > 0$ and $\omega = 0, 1$, we have

$$\mathrm{P}\Big(\frac{1}{n}\sum_{i=1}^{n}\max_{j\in\mathcal{J}_\phi(i)}\|\widehat{U}_{\phi,\omega,j} - \widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,j} + U_{\phi,\omega,i}\| \geq n^{-\frac{1}{2}}\epsilon\Big)$$

$$\leq \mathrm{P}\Big(\sqrt{n}\sup_{x,y\in\mathcal{X},\|\phi_\omega(x)-\phi_\omega(y)\|\leq\delta}\|(\widehat{\phi}_\omega - \phi_\omega)(x) - (\widehat{\phi}_\omega - \phi_\omega)(y)\| \geq \epsilon\Big) + \mathrm{P}\Big(\max_{i\in[\![n]\!]}\max_{j\in\mathcal{J}_\phi(i)}\|U_{\phi,\omega,j} - U_{\phi,\omega,i}\| \geq \delta\Big),$$

which holds for any $\delta > 0$.

Taking $n \to \infty$ and then $\delta \to 0$, by Assumption 5.7 and $M/n \to 0$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\max_{j\in\mathcal{J}_\phi(i)}\|\widehat{U}_{\phi,\omega,j} - \widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,j} + U_{\phi,\omega,i}\| = o_P(n^{-1/2}). \tag{B.5}$$

Plugging (B.3), (B.4), (B.5) into (B.2) and using Assumption 5.6 yields

$$|B_n - \widehat{B}_n|$$

$$\lesssim O_P((M/n)^{k/m} + n^{-k/2}) + \sum_{\ell=1}^{k-1}O_P(n^{-\gamma_\ell}((M/n)^{\ell/m} + n^{-\ell/2})) + O_P((M/n)^{1/m}n^{-1/2}) + o_P(n^{-1/2}).$$

This completes the proof by the selection of M.

B.2 Proof of Theorem A.1

Note that by the multinomial theorem,

$$\mathrm{E}\big[(\widehat{r}_\phi(x))^p\big] = \mathrm{E}\Big[\Big(\frac{N_0}{N_1}\frac{K_\phi(x)}{M}\Big)^p\Big] = \Big(\frac{N_0}{N_1 M}\Big)^p\mathrm{E}\Big[\Big(\sum_{j=1}^{N_1}\mathbb{1}\Big(Z_j \in A_\phi(x)\Big)\Big)^p\Big]$$

$$= \Big(\frac{N_0}{N_1 M}\Big)^p\sum_{p_1+\cdots+p_{N_1}=p;\ p_1,\ldots,p_{N_1}\geq 0}\binom{p}{p_1,\ldots,p_{N_1}}\mathrm{E}\Big[\prod_{j=1}^{N_1}\mathbb{1}\Big(Z_j \in A_\phi(x)\Big)^{p_j}\Big]$$

$$= \Big(\frac{N_0}{N_1 M}\Big)^p\sum_{p_1+\cdots+p_{N_1}=p;\ p_1,\ldots,p_{N_1}\geq 0}\binom{p}{p_1,\ldots,p_{N_1}}\mathrm{P}\Big(Z_j \in A_\phi(x) : p_j > 0\Big).$$

Then by Lemma A.3, we have

$$\lim_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^{\sum_{j=1}^{N_1}\mathbb{1}(p_j>0)}\mathrm{P}\Big(Z_j \in A_\phi(x) : p_j > 0\Big) = \big[r(\phi(x))\big]^{\sum_{j=1}^{N_1}\mathbb{1}(p_j>0)}.$$

Note that for $p_1,\ldots,p_{N_1} \geq 0$ with $p_1+\cdots+p_{N_1} = p$, the number of terms such that $\sum_{j=1}^{N_1}\mathbb{1}(p_j > 0) = k$ is of order N_1^k for any $k \in [\![p]\!]$. Also note that $\binom{p}{p_1,\ldots,p_{N_1}}$ is bounded. Therefore if $MN_1/N_0 \to$

∞, we have

$$\lim_{N_0 \to \infty} \mathrm{E}\big[(\widehat{r}_\phi(x))^p\big] = \lim_{N_0 \to \infty} \frac{1}{N_1^p} \binom{N_1}{p} \binom{p}{1, \ldots, 1} \big[r(\phi(x))\big]^p = \big[r(\phi(x))\big]^p.$$

This completes the proof.

B.3   Proof of Lemma A.3

We only consider those $x \in \mathcal{X}$ such that $f_0(\phi(x)) > 0$ and $\phi(x)$ is a continuous point of f_0 and f_1. We seperate the proof into two cases depending on whether $f_1(\phi(x))$ is zero.

Part I. We first consider the simple case where $p = 1$ and Assumption A.1 holds for p.

Case I. $f_1(\phi(x)) > 0$. Since $\phi(x)$ is a continuous point of f_0 and f_1, for any $\epsilon \in (0, 1)$, there exists some $\delta = \delta_x > 0$ such that for any $z \in \mathcal{X}$ with $\|\phi(z) - \phi(x)\| \leq 3\delta$, we have $|f_0(\phi(z)) - f_0(\phi(x))| \leq \epsilon f_0(\phi(x))$ and $|f_1(\phi(z)) - f_1(\phi(x))| \leq \epsilon f_1(\phi(x))$. Denote the closed ball in $\mathbb{R}^m$ centered at x with radius δ by $B_{x,\delta}$, and the Lebesgue measure by λ. Then for any $z \in \mathcal{X}$ with $\|\phi(z) - \phi(x)\| \leq \delta$, we have

$$\left| \frac{\nu_0(B_{\phi(x),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(x),\|\phi(z)-\phi(x)\|})} - f_0(\phi(x)) \right| \leq \epsilon f_0(\phi(x)), \quad \left| \frac{\nu_0(B_{\phi(z),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(z),\|\phi(z)-\phi(x)\|})} - f_0(\phi(x)) \right| \leq \epsilon f_0(\phi(x)),$$

$$\left| \frac{\nu_1(B_{\phi(x),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(x),\|\phi(z)-\phi(x)\|})} - f_1(\phi(x)) \right| \leq \epsilon f_1(\phi(x)), \quad \left| \frac{\nu_1(B_{\phi(z),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(z),\|\phi(z)-\phi(x)\|})} - f_1(\phi(x)) \right| \leq \epsilon f_1(\phi(x)).$$

Accordingly, if $\|\phi(z) - \phi(x)\| \leq \delta$, we have

$$\frac{1-\epsilon}{1+\epsilon} \frac{f_0(\phi(x))}{f_1(\phi(x))} \leq \frac{\nu_0(B_{\phi(z),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(z),\|\phi(z)-\phi(x)\|})} \frac{\lambda(B_{\phi(x),\|\phi(z)-\phi(x)\|})}{\nu_1(B_{\phi(x),\|\phi(z)-\phi(x)\|})} \leq \frac{1+\epsilon}{1-\epsilon} \frac{f_0(\phi(x))}{f_1(\phi(x))}.$$

Since $\lambda(B_{\phi(z),\|\phi(z)-\phi(x)\|}) = \lambda(B_{\phi(x),\|\phi(z)-\phi(x)\|})$, we then have

$$\frac{1-\epsilon}{1+\epsilon} \frac{f_0(\phi(x))}{f_1(\phi(x))} \leq \frac{\nu_0(B_{\phi(z),\|\phi(z)-\phi(x)\|})}{\nu_1(B_{\phi(x),\|\phi(z)-\phi(x)\|})} \leq \frac{1+\epsilon}{1-\epsilon} \frac{f_0(\phi(x))}{f_1(\phi(x))}.$$

On the other hand, consider any $\epsilon' \in (0, 1)$. For any $z \in \mathcal{X}$ such that $\|\phi(z) - \phi(x)\| > \delta$, as long as ϵ' small enough such that $\epsilon' \mathrm{diam}(\mathcal{X}) < \delta/2$, where $\mathrm{diam}(\mathcal{X})$ is the diameter of $\mathcal{X}$, we have $B_{y,\delta/2} \subset B_{\phi(z),\|\phi(z)-\phi(x)\|-\delta/2} \subset B_{\phi(z),(1-\epsilon')\|\phi(z)-\phi(x)\|}$, where $y \in \mathbb{R}^m$ is taken such that y is the intersection point of the surface of $B_{\phi(x),\delta}$ and the line connecting $\phi(z)$ and $\phi(x)$. Then

$$\nu_0(B_{\phi(z),(1-\epsilon')\|\phi(z)-\phi(x)\|}) \geq \nu_0(B_{y,\delta/2}) \geq (1-\epsilon)f_0(\phi(x))\lambda(B_{y,\delta/2}) = (1-\epsilon)f_0(\phi(x))\lambda(B_{0,\delta/2}).$$

Let $\eta_N = 4\log(N_0/M)$. Since $M\log N_0/N_0 \to 0$, we can take N_0 large enough so that

$$\eta_N \frac{M}{N_0} = 4\frac{M}{N_0}\log\left(\frac{N_0}{M}\right) < (1-\epsilon)f_0(\phi(x))\lambda(B_{0,\delta/2}).$$

Then for any $z \in \mathcal{X}$ such that $\nu_0(B_{\phi(z),(1-\epsilon')\|\phi(z)-\phi(x)\|}) \leq \eta_N M/N_0$, we have $\|\phi(z) - \phi(x)\| < \delta$ since otherwise it would contradict the selection of η_N.

Upper bound. Let $\Phi_M(z)$ be the M-th order statistics of $\{\|\phi(X_i) - \phi(z)\|\}_{i=1}^{N_0}$. By the definition of $A_\phi(x)$, we have for any $\epsilon' \in (0, 1)$,

$$\mathrm{P}\Big(Z_1 \in A_\phi(x)\Big) = \mathrm{P}\Big(\|\widehat{\phi}(x) - \widehat{\phi}(Z_1)\| \leq \widehat{\Phi}_M(Z_1)\Big)$$

$$\leq \mathrm{P}\Big(\|\phi(x) - \phi(Z_1)\| - 2\|\widehat{\phi} - \phi\|_\infty \leq \Phi_M(Z_1) + 2\|\widehat{\phi} - \phi\|_\infty\Big)$$

$$= \mathrm{P}\Big(\|\phi(x) - \phi(Z_1)\| - 4\|\widehat{\phi} - \phi\|_\infty \leq \Phi_M(Z_1), 4\|\widehat{\phi} - \phi\|_\infty \leq \epsilon'\|\phi(x) - \phi(Z_1)\|\Big)$$

$$+ \mathrm{P}\Big(\|\phi(x) - \phi(Z_1)\| - 4\|\widehat{\phi} - \phi\|_\infty \leq \Phi_M(Z_1), 4\|\widehat{\phi} - \phi\|_\infty > \epsilon'\|\phi(x) - \phi(Z_1)\|\Big). \qquad \text{(B.6)}$$

For the first term in (B.6), note that $[\phi(X_i)]_{i=1}^{N_0}$ are i.i.d from ν_0, and then $\nu_0(B_{\phi(Z_1),\|\phi(X_i)-\phi(Z_1)\|})$ are i.i.d from $U(0,1)$ and are independent of Z_1 by the probability integral transform. Then

$$\mathrm{P}\Big(\|\phi(x) - \phi(Z_1)\| - 4\|\widehat{\phi} - \phi\|_\infty \leq \Phi_M(Z_1), 4\|\widehat{\phi} - \phi\|_\infty \leq \epsilon'\|\phi(x) - \phi(Z_1)\|\Big)$$

$$\leq \mathrm{P}\Big((1-\epsilon')\|\phi(x) - \phi(Z_1)\| \leq \Phi_M(Z_1)\Big)$$

$$= \mathrm{P}\Big(\nu_0(B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)})\Big)$$

$$\leq \mathrm{P}\Big(\nu_0(B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}) \leq \eta_N \frac{M}{N_0}\Big) + \mathrm{P}\Big(U_{(M)} > \eta_N \frac{M}{N_0}\Big),$$

where $U_{(M)}$ is the M-th order statistic of N_0 independent random variables from $U(0,1)$.

By the selection of η_N, and taking ϵ' small and N_0 large enough, we have

$$\mathrm{P}\Big(\nu_0(B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}) \leq \eta_N \frac{M}{N_0}\Big)$$

$$\leq \mathrm{P}\Big(\nu_0(B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}), \|\phi(x) - \phi(Z_1)\| \leq \delta\Big).$$

Under the event $\{\|\phi(x) - \phi(Z_1)\| \leq \delta\}$, we have

$$\nu_0(B_{\phi(Z_1),\|\phi(x)-\phi(Z_1)\|}) - \nu_0(B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|})$$

$$= \int_{B_{\phi(Z_1),\|\phi(x)-\phi(Z_1)\|} \setminus B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|}} f_0(y)\mathrm{d}y$$

$$\leq (1+\epsilon)f_0(\phi(x))\lambda(B_{\phi(Z_1),\|\phi(x)-\phi(Z_1)\|} \setminus B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|})$$

$$= (1+\epsilon)f_0(\phi(x))V_m[1 - (1-\epsilon')^d]\|\phi(x) - \phi(Z_1)\|^d$$

$$\leq (1+\epsilon)f_0(\phi(x))V_m d\epsilon'\|\phi(x) - \phi(Z_1)\|^d$$

$$= (1+\epsilon)f_0(\phi(x))d\epsilon'\lambda(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|})$$

$$\leq \frac{(1+\epsilon)f_0(\phi(x))d\epsilon'}{(1-\epsilon)f_1(\phi(x))}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}),$$

where V_m is the Lebesgue measure of the $m-$dimensional unit ball, and

$$\nu_0(B_{\phi(Z_1),\|\phi(x)-\phi(Z_1)\|}) \geq \frac{(1-\epsilon)f_0(\phi(x))}{(1+\epsilon)f_1(\phi(x))}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}).$$

From the probability integral transform, we have $\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|})$ is from $U(0,1)$ and then for $U \sim U(0,1)$,

$$\mathrm{P}\Big(\nu_0(B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}), \|\phi(x) - \phi(Z_1)\| \leq \delta\Big)$$

$$\leq \mathrm{P}\Big(\Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)\frac{f_0(\phi(x))}{f_1(\phi(x))}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)})\Big)$$

$$= \mathrm{P}\Big(\Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)\frac{f_0(\phi(x))}{f_1(\phi(x))}U \le U_{(M)}\Big).$$

We can check that

$$\lim_{N_0 \to \infty} \frac{N_0}{M}\mathrm{P}\Big(\Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)\frac{f_0(\phi(x))}{f_1(\phi(x))}U \le U_{(M)}\Big) = \Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}.$$

Note that $\eta_N \to \infty$ as $N_0 \to \infty$ since $M/N_0 \to 0$. Then from the Chernoff bound and for N_0 sufficiently large, we have

$$\frac{N_0}{M}\mathrm{P}\Big(U_{(M)} > \eta_N \frac{M}{N_0}\Big) = \frac{N_0}{M}\mathrm{P}\Big(\mathrm{Bin}\Big(N_0, \eta_N \frac{M}{N_0}\Big) \le M\Big)$$

$$\le \frac{N_0}{M}\exp\Big((1 + \log \eta_N - \eta_N)M\Big) \le \frac{N_0}{M}\exp\Big(-\frac{1}{2}\eta_N M\Big) = \Big(\frac{N_0}{M}\Big)^{1-2M}.$$

Since $M/N_0 \to 0$ and $M \ge 1$, we then obtain

$$\lim_{N_0 \to \infty} \frac{N_0}{M}\mathrm{P}\Big(U_{(M)} > \eta_N \frac{M}{N_0}\Big) = 0.$$

Then we obtain

$$\limsup_{N_0 \to \infty} \frac{N_0}{M}\mathrm{P}\Big(\|\phi(x) - \phi(Z_1)\| - 4\|\widehat{\phi} - \phi\|_\infty \le \Phi_M(Z_1), 4\|\widehat{\phi} - \phi\|_\infty \le \epsilon'\|\phi(x) - \phi(Z_1)\|\Big)$$

$$\le \Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}.$$

For the second term in (B.6), we have for any $\delta > 0$,

$$\mathrm{P}\Big(\|\phi(x) - \phi(Z_1)\| - 4\|\widehat{\phi} - \phi\|_\infty \le \Phi_M(Z_1), 4\|\widehat{\phi} - \phi\|_\infty > \epsilon'\|\phi(x) - \phi(Z_1)\|\Big)$$

$$\le \mathrm{P}\Big(\|\phi(x) - \phi(Z_1)\| < 4\|\widehat{\phi} - \phi\|_\infty/\epsilon'\Big)$$

$$\le \mathrm{P}\Big(\|\phi(x) - \phi(Z_1)\| < 4\|\widehat{\phi} - \phi\|_\infty/\epsilon' \le \delta\Big) + \mathrm{P}\Big(4\|\widehat{\phi} - \phi\|_\infty/\epsilon' > \delta\Big).$$

Note that $\widehat{\phi}$ may depend on Z_1. Recall that $\widehat{\phi}_{Z_1 \to z}$ is the estimator of ϕ replacing Z_1 by some $z \in \mathcal{Z}$. Then $\sup_{z \in \mathcal{Z}}\|\widehat{\phi}_{Z_1 \to z} - \phi\|_\infty$ is independent of Z_1 and then we have

$$\mathrm{P}\Big(\|\phi(x) - \phi(Z_1)\| < 4\|\widehat{\phi} - \phi\|_\infty/\epsilon' \le \delta\Big)$$

$$\le \mathrm{P}\Big(\lambda(B_{\phi(x), \|\phi(x) - \phi(Z_1)\|}) \le V_m(4/\epsilon')^d\|\widehat{\phi} - \phi\|_\infty^d, \|\phi(x) - \phi(Z_1)\| \le \delta\Big)$$

$$\le \mathrm{P}\Big(\nu_1(B_{\phi(x), \|\phi(x) - \phi(Z_1)\|}) \le (1 + \epsilon)f_1(\phi(x))V_m(4/\epsilon')^d\|\widehat{\phi} - \phi\|_\infty^d\Big)$$

$$\le \mathrm{P}\Big(\nu_1(B_{\phi(x), \|\phi(x) - \phi(Z_1)\|}) \le (1 + \epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z \in \mathcal{Z}}\|\widehat{\phi}_{Z_1 \to z} - \phi\|_\infty^d\Big)$$

$$= \mathrm{E}\{[(1 + \epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z \in \mathcal{Z}}\|\widehat{\phi}_{Z_1 \to z} - \phi\|_\infty^d] \wedge 1\}.$$

By the Markov inequality,

$$\mathrm{P}\Big(4\|\widehat{\phi} - \phi\|_\infty/\epsilon' > \delta\Big) \le \delta^{-d}(4/\epsilon')^d\mathrm{E}[\|\widehat{\phi} - \phi\|_\infty^d] \le \delta^{-d}(4/\epsilon')^d\mathrm{E}[\sup_{z \in \mathcal{Z}}\|\widehat{\phi}_{Z_1 \to z} - \phi\|_\infty^d].$$

By Assumption A.1, we have

$$\lim_{N_0 \to \infty} \frac{N_0}{M}\mathrm{P}\Big(\|\phi(x) - \phi(Z_1)\| - 4\|\widehat{\phi} - \phi\|_\infty \le \Phi_M(Z_1), 4\|\widehat{\phi} - \phi\|_\infty > \epsilon'\|\phi(x) - \phi(Z_1)\|\Big) = 0.$$

By (B.6) and ϵ, ϵ' are arbitrary, we obtain

$$\limsup_{N_0 \to \infty} \frac{N_0}{M} \mathrm{P}\Big(Z_1 \in A_\phi(x)\Big) \leq \frac{f_1(\phi(x))}{f_0(\phi(x))}.$$

Lower bound. For any $\epsilon' \in (0, 1)$, we have

$$\mathrm{P}\Big(Z_1 \in A_\phi(x)\Big) = \mathrm{P}\Big(\|\widehat{\phi}(x) - \widehat{\phi}(Z_1)\| \leq \widehat{\Phi}_M(Z_1)\Big)$$

$$\geq \mathrm{P}\Big(\|\phi(x) - \phi(Z_1)\| + 2\|\widehat{\phi} - \phi\|_\infty \leq \Phi_M(Z_1) - 2\|\widehat{\phi} - \phi\|_\infty\Big)$$

$$\geq \mathrm{P}\Big(\|\phi(x) - \phi(Z_1)\| + 4\|\widehat{\phi} - \phi\|_\infty \leq \Phi_M(Z_1), 4\|\widehat{\phi} - \phi\|_\infty \leq \epsilon'\|\phi(x) - \phi(Z_1)\|\Big)$$

$$\geq \mathrm{P}\Big((1 + \epsilon')\|\phi(x) - \phi(Z_1)\| \leq \Phi_M(Z_1), 4\|\widehat{\phi} - \phi\|_\infty \leq \epsilon'\|\phi(x) - \phi(Z_1)\|\Big)$$

$$\geq \mathrm{P}\Big((1 + \epsilon')\|\phi(x) - \phi(Z_1)\| \leq \Phi_M(Z_1)\Big) - \mathrm{P}\Big(4\|\widehat{\phi} - \phi\|_\infty > \epsilon'\|\phi(x) - \phi(Z_1)\|\Big)$$

$$= \mathrm{P}\Big(\nu_0(B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)})\Big) - \mathrm{P}\Big(4\|\widehat{\phi} - \phi\|_\infty > \epsilon'\|\phi(x) - \phi(Z_1)\|\Big)$$

$$\geq \mathrm{P}\Big(\nu_0(B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}) \leq \eta_N \frac{M}{N_0}\Big) - \mathrm{P}\Big(4\|\widehat{\phi} - \phi\|_\infty > \epsilon'\|\phi(x) - \phi(Z_1)\|\Big).$$

Note that by the selection of η_N, and taking ϵ' small and N_0 large enough, we have

$$\mathrm{P}\Big(\nu_0(B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}) \leq \eta_N \frac{M}{N_0}\Big)$$

$$= \mathrm{P}\Big(\nu_0(B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}) \leq \eta_N \frac{M}{N_0}, \|\phi(x) - \phi(Z_1)\| \leq \delta\Big).$$

Under the event $\{\|\phi(x) - \phi(Z_1)\| \leq \delta\}$, we have $B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|} \subset B_{\phi(x),3\delta}$, and then

$$\nu_0(B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|}) - \nu_0(B_{\phi(Z_1),\|\phi(x)-\phi(Z_1)\|})$$

$$= \int_{B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|} \setminus B_{\phi(Z_1),\|\phi(x)-\phi(Z_1)\|}} f_0(y)\mathrm{d}y$$

$$\leq (1 + \epsilon)f_0(\phi(x))\lambda(B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|} \setminus B_{\phi(Z_1),\|\phi(x)-\phi(Z_1)\|})$$

$$= (1 + \epsilon)f_0(\phi(x))V_m[(1 + \epsilon')^d - 1]\|\phi(x) - \phi(Z_1)\|^d$$

$$\leq (1 + \epsilon)f_0(\phi(x))V_m d\epsilon'(1 + \epsilon')^{d-1}\|\phi(x) - \phi(Z_1)\|^d$$

$$= (1 + \epsilon)f_0(\phi(x))d\epsilon'(1 + \epsilon')^{d-1}\lambda(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|})$$

$$\leq \frac{(1 + \epsilon)f_0(\phi(x))d\epsilon'(1 + \epsilon')^{d-1}}{(1 - \epsilon)f_1(\phi(x))}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}),$$

and

$$\nu_0(B_{\phi(Z_1),\|\phi(x)-\phi(Z_1)\|}) \leq \frac{(1 + \epsilon)f_0(\phi(x))}{(1 - \epsilon)f_1(\phi(x))}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}).$$

Then

$$\mathrm{P}\Big(\nu_0(B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}) \leq \eta_N \frac{M}{N_0}, \|\phi(x) - \phi(Z_1)\| \leq \delta\Big)$$

$$\geq \mathrm{P}\Big(\frac{(1 + \epsilon)f_0(\phi(x))}{(1 - \epsilon)f_1(\phi(x))}\Big(1 + d\epsilon'(1 + \epsilon')^{d-1}\Big)\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}) \leq \eta_N \frac{M}{N_0}, \|\phi(x) - \phi(Z_1)\| \leq \delta\Big)$$

$$= \mathrm{P}\Big(\frac{(1+\epsilon)f_0(\phi(x))}{(1-\epsilon)f_1(\phi(x))}\Big(1+d\epsilon'(1+\epsilon')^{d-1}\Big)\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}) \leq \eta_N\frac{M}{N_0}\Big)$$

$$\geq \mathrm{P}\Big(\frac{(1+\epsilon)f_0(\phi(x))}{(1-\epsilon)f_1(\phi(x))}\Big(1+d\epsilon'(1+\epsilon')^{d-1}\Big)\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)})\Big) - \mathrm{P}\Big(U_{(M)} > \eta_N\frac{M}{N_0}\Big)$$

$$= \mathrm{P}\Big(\frac{(1+\epsilon)f_0(\phi(x))}{(1-\epsilon)f_1(\phi(x))}\Big(1+d\epsilon'(1+\epsilon')^{d-1}\Big)U \leq U_{(M)}\Big) - \mathrm{P}\Big(U_{(M)} > \eta_N\frac{M}{N_0}\Big).$$

The second last equality is from the fact that for $z \in \mathcal{X}$ such that $\|\phi(z) - \phi(x)\| > \delta$,

$$\frac{(1+\epsilon)f_0(\phi(x))}{(1-\epsilon)f_1(\phi(x))}\Big(1+d\epsilon'(1+\epsilon')^{d-1}\Big)\nu_1(B_{\phi(x),\|\phi(x)-\phi(z)\|}) \geq \frac{(1+\epsilon)f_0(\phi(x))}{(1-\epsilon)f_1(\phi(x))}\nu_1(B_{\phi(x),\delta})$$

$$\geq \frac{(1+\epsilon)f_0(\phi(x))}{(1-\epsilon)f_1(\phi(x))}f_1(\phi(x))(1-\epsilon)\lambda(B_{0,\delta}) > \eta_N\frac{M}{N_0}$$

by the selection of η_N.

We can check that

$$\lim_{N_0\to\infty}\frac{N_0}{M}\mathrm{P}\Big(\frac{(1+\epsilon)f_0(\phi(x))}{(1-\epsilon)f_1(\phi(x))}\Big(1+d\epsilon'(1+\epsilon')^{d-1}\Big)U \leq U_{(M)}\Big) = \frac{1-\epsilon}{1+\epsilon}\Big(1+d\epsilon'(1+\epsilon')^{d-1}\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}.$$

By ϵ, ϵ' are arbitrary, we obtain

$$\liminf_{N_0\to\infty}\frac{N_0}{M}\mathrm{P}\Big(Z_1 \in A_\phi(x)\Big) \geq \frac{f_1(\phi(x))}{f_0(\phi(x))}.$$

Combining the upper bound and the lower bound yields

$$\lim_{N_0\to\infty}\frac{N_0}{M}\mathrm{P}\Big(Z_1 \in A_\phi(x)\Big) = \frac{f_1(\phi(x))}{f_0(\phi(x))}.$$

Case II. $f_1(\phi(x)) = 0$.

For any $\epsilon \in (0,1)$, there exists some $\delta = \delta_x > 0$ such that for any $z \in \mathcal{X}$ with $\|\phi(z)-\phi(x)\| \leq 3\delta$, we have $|f_0(\phi(z))-f_0(\phi(x))| \leq \epsilon f_0(\phi(x))$ and $f_1(\phi(z)) \leq \epsilon$. Then for any $z \in \mathcal{X}$ with $\|\phi(z)-\phi(x)\| \leq \delta$, we have

$$\Big|\frac{\nu_0(B_{\phi(z),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(z),\|\phi(z)-\phi(x)\|})} - f_0(\phi(x))\Big| \leq \epsilon f_0(\phi(x)), \quad \Big|\frac{\nu_1(B_{\phi(x),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(x),\|\phi(z)-\phi(x)\|})}\Big| \leq \epsilon.$$

We consider the same decomposition as (B.6). For the first term in (B.6), we still have

$$\mathrm{P}\Big(\|\phi(x)-\phi(Z_1)\| - 4\|\widehat{\phi}-\phi\|_\infty \leq \Phi_M(Z_1), 4\|\widehat{\phi}-\phi\|_\infty \leq \epsilon'\|\phi(x)-\phi(Z_1)\|\Big)$$

$$\leq \mathrm{P}\Big(\nu_0(B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}), \|\phi(x)-\phi(Z_1)\| \leq \delta\Big) + \mathrm{P}\Big(U_{(M)} > \eta_N\frac{M}{N_0}\Big).$$

Note that for any $z \in \mathcal{X}$ with $\|\phi(x)-\phi(z)\| \leq \delta$,

$$\Big|\frac{\nu_0(B_{\phi(z),(1-\epsilon')\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(z),(1-\epsilon')\|\phi(z)-\phi(x)\|})} - f_0(\phi(x))\Big| \leq \epsilon f_0(\phi(x)),$$

and

$$\frac{\nu_1(B_{\phi(x),\|\phi(x)-\phi(z)\|})}{\lambda(B_{\phi(z),(1-\epsilon')\|\phi(z)-\phi(x)\|})} = \frac{\lambda(B_{\phi(x),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(z),(1-\epsilon')\|\phi(z)-\phi(x)\|})}\frac{\nu_1(B_{\phi(x),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(x),\|\phi(z)-\phi(x)\|})}$$

$$\leq \frac{\lambda(B_{\phi(x),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(z),(1-\epsilon')\|\phi(z)-\phi(x)\|})}\epsilon = (1-\epsilon')^{-d}\epsilon.$$

Then
$$\mathrm{P}\Big(\nu_0(B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|}) \le \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}), \|\phi(x)-\phi(Z_1)\| \le \delta\Big)$$
$$\le \mathrm{P}\Big((1-\epsilon')^d\epsilon^{-1}(1-\epsilon)f_0(\phi(x))\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \le \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)})\Big)$$
$$= \mathrm{P}\Big((1-\epsilon')^d\epsilon^{-1}(1-\epsilon)f_0(\phi(x))U \le U_{(M)}\Big).$$

We can check that
$$\lim_{N_0\to\infty}\frac{N_0}{M}\mathrm{P}\Big((1-\epsilon')^d\epsilon^{-1}(1-\epsilon)f_0(\phi(x))U \le U_{(M)}\Big) = \epsilon(1-\epsilon')^{-d}(1-\epsilon)^{-1}\frac{1}{f_0(\phi(x))}.$$

Then we obtain
$$\limsup_{N_0\to\infty}\frac{N_0}{M}\mathrm{P}\Big(\|\phi(x)-\phi(Z_1)\| - 4\|\widehat\phi-\phi\|_\infty \le \Phi_M(Z_1), 4\|\widehat\phi-\phi\|_\infty \le \epsilon'\|\phi(x)-\phi(Z_1)\|\Big)$$
$$\le \epsilon(1-\epsilon')^{-d}(1-\epsilon)^{-1}\frac{1}{f_0(\phi(x))}.$$

For the second term in (B.6), we still have
$$\mathrm{P}\Big(\|\phi(x)-\phi(Z_1)\| - 4\|\widehat\phi-\phi\|_\infty \le \Phi_M(Z_1), 4\|\widehat\phi-\phi\|_\infty > \epsilon'\|\phi(x)-\phi(Z_1)\|\Big)$$
$$\le \mathrm{P}\Big(\|\phi(x)-\phi(Z_1)\| < 4\|\widehat\phi-\phi\|_\infty/\epsilon' \le \delta\Big) + \mathrm{P}\Big(4\|\widehat\phi-\phi\|_\infty/\epsilon' > \delta\Big).$$

Note that
$$\mathrm{P}\Big(\|\phi(x)-\phi(Z_1)\| < 4\|\widehat\phi-\phi\|_\infty/\epsilon' \le \delta\Big)$$
$$\le \mathrm{P}\Big(\lambda(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \le V_m(4/\epsilon')^d\|\widehat\phi-\phi\|_\infty^d, \|\phi(x)-\phi(Z_1)\| \le \delta\Big)$$
$$\le \mathrm{P}\Big(\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \le \epsilon V_m(4/\epsilon')^d\|\widehat\phi-\phi\|_\infty^d\Big)$$
$$\le \mathrm{P}\Big(\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \le \epsilon V_m(4/\epsilon')^d\sup_{z\in\mathcal{Z}}\|\widehat\phi_{Z_1\to z}-\phi\|_\infty^d\Big)$$
$$= \mathrm{E}\{[\epsilon V_m(4/\epsilon')^d\sup_{z\in\mathcal{Z}}\|\widehat\phi_{Z_1\to z}-\phi\|_\infty^d] \wedge 1\}.$$

By ϵ, ϵ' are arbitrary, we obtain
$$\lim_{N_0\to\infty}\frac{N_0}{M}\mathrm{P}\Big(Z_1 \in A_\phi(x)\Big) = 0 = \frac{f_1(\phi(x))}{f_0(\phi(x))}.$$

Part II. We then consider the general case where p is a fixed positive integer and Assumption A.1 holds for p. We only consider the case where $f_1(\phi(x)) > 0$. The case with $f_1(\phi(x)) = 0$ can be established in a similar way.

Let $\eta_N = \eta_{N,p} = 4p\log(N_0/M)$. We also take N_0 sufficiently large so that
$$\eta_N\frac{M}{N_0} = 4p\frac{M}{N_0}\log\Big(\frac{N_0}{M}\Big) < (1-\epsilon)f_0(\phi(x))\lambda(B_{0,\delta/2}).$$

Then
$$\mathrm{P}\Big(Z_1,\ldots,Z_p \in A_\phi(x)\Big) = \mathrm{P}\Big(\|\widehat\phi(x)-\widehat\phi(Z_k)\| \le \widehat\Phi_M(Z_k), \forall k \in [\![p]\!]\Big)$$
$$\le \mathrm{P}\Big(\|\phi(x)-\phi(Z_k)\| - 2\|\widehat\phi-\phi\|_\infty \le \Phi_M(Z_k) + 2\|\widehat\phi-\phi\|_\infty, \forall k \in [\![p]\!]\Big)$$

36

$$= \sum_{S \subset [\![p]\!]} \mathrm{P}\Big(\|\phi(x) - \phi(Z_k)\| - 4\|\widehat{\phi} - \phi\|_\infty \leq \Phi_M(Z_k), 4\|\widehat{\phi} - \phi\|_\infty \leq \epsilon'\|\phi(x) - \phi(Z_k)\| \text{ for } k \in S,$$

$$4\|\widehat{\phi} - \phi\|_\infty > \epsilon'\|\phi(x) - \phi(Z_k)\| \text{ for } k \notin S\Big)$$

$$\leq \sum_{S \subset [\![p]\!]} \mathrm{P}\Big(\nu_0(B_{\phi(Z_k),(1-\epsilon')\|\phi(x)-\phi(Z_k)\|}) \leq \nu_0(B_{\phi(Z_k),\Phi_M(Z_k)}) \text{ for } k \in S, 4\|\widehat{\phi} - \phi\|_\infty > \epsilon' \max_{k\notin S}\|\phi(x) - \phi(Z_k)\|\Big).$$

$$(\text{B.7})$$

Let $W_k = \nu_0(B_{\phi(Z_k),(1-\epsilon')\|\phi(x)-\phi(Z_k)\|})$ and $V_k = \nu_0(B_{\phi(Z_k),\Phi_M(Z_k)})$ for any $k \in [\![p]\!]$. Then $[W_k]_{k=1}^p$ are i.i.d. since $[Z_k]_{k=1}^p$ are i.i.d.. For any $k \in [\![p]\!]$ and $Z_k \in \mathcal{X}$ given, $V_k \mid Z_k$ has the same distribution as $U_{(M)}$. Then for any $k \in [\![p]\!]$, V_k has the same distribution as $U_{(M)}$, and V_k is independent of Z_k.

Fix $S \subset [\![p]\!]$. Let $W_{\max} = \max_{k\in S} W_k$ and $V_{\max} = \max_{k\in S} V_k$. Then

$$\mathrm{P}\Big(\nu_0(B_{\phi(Z_k),(1-\epsilon')\|\phi(x)-\phi(Z_k)\|}) \leq \nu_0(B_{\phi(Z_k),\Phi_M(Z_k)}) \text{ for } k \in S, 4\|\widehat{\phi} - \phi\|_\infty > \epsilon' \max_{k\notin S}\|\phi(x) - \phi(Z_k)\|\Big)$$

$$\leq \mathrm{P}\Big(W_{\max} < V_{\max}, 4\|\widehat{\phi} - \phi\|_\infty > \epsilon' \max_{k\notin S}\|\phi(x) - \phi(Z_k)\|\Big)$$

$$\leq \mathrm{P}\Big(W_{\max} < V_{\max} \leq \eta_N \frac{M}{N_0}, \max_{k\notin S}\|\phi(x) - \phi(Z_k)\| < 4\|\widehat{\phi} - \phi\|_\infty/\epsilon' \leq \delta\Big)$$

$$+ \mathrm{P}\Big(V_{\max} > \eta_N \frac{M}{N_0}\Big) + \mathrm{P}\Big(4\|\widehat{\phi} - \phi\|_\infty/\epsilon' > \delta\Big). \qquad (\text{B.8})$$

For the first term in (B.8), by the selection of η_N, and taking $\epsilon' < 1/2$ and N_0 large enough, we have

$$\mathrm{P}\Big(W_{\max} < V_{\max} \leq \eta_N \frac{M}{N_0}, \max_{k\notin S}\|\phi(x) - \phi(Z_k)\| < 4\|\widehat{\phi} - \phi\|_\infty/\epsilon' \leq \delta\Big)$$

$$\leq \mathrm{P}\Big(W_{\max} < V_{\max} \leq \eta_N \frac{M}{N_0}, \max_{k\notin S}\|\phi(x) - \phi(Z_k)\| < 4\|\widehat{\phi} - \phi\|_\infty/\epsilon' \leq \delta, \max_{k\in S}\|\phi(x) - \phi(Z_k)\| \leq \delta\Big).$$

Let $W'_k = \nu_0(B_{\phi(Z_k),\|\phi(x)-\phi(Z_k)\|})$ and $W'_{\max} = \max_{k\in S} W'_k$. Under the event $\{\max_{k\in S}\|\phi(x) - \phi(Z_k)\| \leq \delta\}$, we have

$$W'_{\max} - W_{\max} \leq \max_{k\in S}[\nu_0(B_{\phi(Z_k),\|\phi(x)-\phi(Z_k)\|}) - \nu_0(B_{\phi(Z_k),(1-\epsilon')\|\phi(x)-\phi(Z_k)\|})]$$

$$\leq \frac{(1+\epsilon)f_0(\phi(x))d\epsilon'}{(1-\epsilon)f_1(\phi(x))} \max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}).$$

and

$$W'_{\max} \geq \frac{(1-\epsilon)f_0(\phi(x))}{(1+\epsilon)f_1(\phi(x))} \max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}).$$

On the other hand, recall that $\widehat{\phi}_{(Z_1,\ldots,Z_p)\to z}$ is the estimator replacing $(Z_1,\ldots,Z_p)$ by z for $z \in \mathcal{Z}^p$. Then $\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\ldots,Z_p)\to z} - \phi\|_\infty$ is independent of $(Z_1,\ldots,Z_p)$. Note that

$$\max_{k\notin S}\|\phi(x) - \phi(Z_k)\| < 4\|\widehat{\phi} - \phi\|_\infty/\epsilon' \leq \delta$$

implies that

$$\max_{k\notin S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}) \leq (1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\ldots,Z_p)\to z} - \phi\|_\infty^d.$$

37

Then
$$
\mathrm{P}\Big(W_{\max} < V_{\max} \le \eta_N \frac{M}{N_0}, \max_{k\notin S}\|\phi(x)-\phi(Z_k)\| < 4\|\widehat{\phi}-\phi\|_\infty/\epsilon' \le \delta, \max_{k\in S}\|\phi(x)-\phi(Z_k)\| \le \delta\Big)
$$

$$
\le \mathrm{P}\Big(\Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)\frac{f_0(\phi(x))}{f_1(\phi(x))}\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}) < V_{\max},
$$

$$
\max_{k\notin S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}) \le (1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d\Big)
$$

$$
= \mathrm{E}\Big[\mathbb{1}\Big(\Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)\frac{f_0(\phi(x))}{f_1(\phi(x))}\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}) < V_{\max}\Big)
$$

$$
\Big([(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d] \wedge 1\Big)^{p-|S|}\Big],
$$

since $\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty$, $V_{\max}$ and $[Z_k]_{k\in S}$ are all independent with $[Z_k]_{k\notin S}$.

Note that
$$
\mathrm{E}\Big[\mathbb{1}\Big(\Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)\frac{f_0(\phi(x))}{f_1(\phi(x))}\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}) < V_{\max}\Big)
$$

$$
\Big([(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d] \wedge 1\Big)^{p-|S|}\Big]
$$

$$
= \int_0^1 |S|u^{|S|-1}\mathrm{E}\Big[\mathbb{1}\Big(\Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)\frac{f_0(\phi(x))}{f_1(\phi(x))}u < V_{\max}\Big)
$$

$$
\Big([(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d] \wedge 1\Big)^{p-|S|}\ \Big|\ \max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}) = u\Big]\mathrm{d}u
$$

$$
= |S|\Big[\Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\frac{M}{N_0}\Big]^{|S|} \int_0^{(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon')\frac{f_0(\phi(x))}{f_1(\phi(x))}\frac{N_0}{M}} u^{|S|-1}\mathrm{E}\Big[\mathbb{1}\Big(V_{\max} > \frac{M}{N_0}u\Big)
$$

$$
\Big([(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d] \wedge 1\Big)^{p-|S|}\ \Big|
$$

$$
\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}) = \Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\frac{M}{N_0}u\Big]\mathrm{d}u. \tag{B.9}
$$

We split the above integral into two parts using 1. For the first part, note that $\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty$ is independent with $\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|})$. Then

$$
\Big(\frac{N_0}{M}\Big)^{p-|S|}\int_0^1 u^{|S|-1}\mathrm{E}\Big[\mathbb{1}\Big(V_{\max} > \frac{M}{N_0}u\Big)\Big([(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d] \wedge 1\Big)^{p-|S|}\ \Big|
$$

$$
\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}) = \Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\frac{M}{N_0}u\Big]\mathrm{d}u
$$

$$
\le \Big(\frac{N_0}{M}\Big)^{p-|S|}\int_0^1 u^{|S|-1}\mathrm{E}\Big[\Big([(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d] \wedge 1\Big)^{p-|S|}\Big]\mathrm{d}u.
$$

If $|S| = p$, we have
$$
\Big(\frac{N_0}{M}\Big)^{p-|S|}\int_0^1 u^{|S|-1}\mathrm{E}\Big[\Big([(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d] \wedge 1\Big)^{p-|S|}\Big]\mathrm{d}u
$$

$$
= \int_0^1 u^{p-1}\mathrm{d}u = \frac{1}{p}. \tag{B.10}
$$

38

If $|S| < p$, by Assumption A.1, we have

$$\limsup_{N_0 \to \infty} \Big(\frac{N_0}{M}\Big)^{p-|S|} \int_0^1 u^{|S|-1} \mathrm{E}\Big[\Big([(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z \in \mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d] \wedge 1\Big)^{p-|S|}\Big]\mathrm{d}u$$

$$\lesssim \limsup_{N_0 \to \infty} \mathrm{E}\Big[\Big(\frac{N_0}{M}\sup_{z \in \mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d\Big)^{p-|S|}\Big] = 0. \tag{B.11}$$

For the second part, we have

$$\int_1^{(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon')\frac{f_0(\phi(x))}{f_1(\phi(x))}\frac{N_0}{M}} u^{|S|-1}\mathrm{E}\Big[\mathbb{1}\Big(V_{\max} > \frac{M}{N_0}u\Big)\Big([(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z \in \mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d] \wedge 1\Big)^{p-|S|}\Big|$$

$$\max_{k \in S} \nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}) = \Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\frac{M}{N_0}u\Big]\mathrm{d}u$$

$$\leq \int_1^\infty u^{|S|-1}\mathrm{E}\Big[\mathbb{1}\Big(V_{\max} > \frac{M}{N_0}u\Big)\Big([(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z \in \mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d] \wedge 1\Big)^{p-|S|}\Big|$$

$$\max_{k \in S} \nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}) = \Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\frac{M}{N_0}u\Big]\mathrm{d}u$$

$$\leq \sum_{k \in S} \int_1^\infty u^{|S|-1}\mathrm{E}\Big[\mathbb{1}\Big(V_k > \frac{M}{N_0}u\Big)\Big([(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z \in \mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d] \wedge 1\Big)^{p-|S|}\Big]\mathrm{d}u, \tag{B.12}$$

where the last step is from the fact that V_k and $\sup_{z \in \mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty$ are independent of $[Z_k]_{k \in S}$ for any $k \in S$.

For any $k \in S$, by the Hölder inequality,

$$\Big(\frac{N_0}{M}\Big)^{p-|S|} \int_1^\infty u^{|S|-1}\mathrm{E}\Big[\mathbb{1}\Big(V_k > \frac{M}{N_0}u\Big)\Big([(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z \in \mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d] \wedge 1\Big)^{p-|S|}\Big]\mathrm{d}u$$

$$\lesssim \int_1^\infty u^{|S|-1}\mathrm{E}\Big[\mathbb{1}\Big(V_k > \frac{M}{N_0}u\Big)\Big(\frac{N_0}{M}\sup_{z \in \mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d\Big)^{p-|S|}\Big]\mathrm{d}u$$

$$\leq \int_1^\infty u^{|S|-1}\Big\{\mathrm{P}\Big(V_k > \frac{M}{N_0}u\Big)\Big\}^{\frac{|S|}{p}}\Big\{\mathrm{E}\Big[\Big(\frac{N_0}{M}\sup_{z \in \mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d\Big)^p\Big]\Big\}^{\frac{p-|S|}{p}}\mathrm{d}u$$

$$= \Big\{\mathrm{E}\Big[\Big(\frac{N_0}{M}\sup_{z \in \mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^d\Big)^p\Big]\Big\}^{\frac{p-|S|}{p}} \int_1^\infty u^{|S|-1}\Big\{\mathrm{P}\Big(V_k > \frac{M}{N_0}u\Big)\Big\}^{\frac{|S|}{p}}\mathrm{d}u.$$

Using the Chernoff bound,

$$\int_1^\infty u^{|S|-1}\Big\{\mathrm{P}\Big(V_k > \frac{M}{N_0}u\Big)\Big\}^{\frac{|S|}{p}}\mathrm{d}u = \int_0^\infty (1+u)^{|S|-1}\Big\{\mathrm{P}\Big(U_{(M)} > \frac{M}{N_0}(1+u)\Big)\Big\}^{\frac{|S|}{p}}\mathrm{d}u$$

$$\leq \int_0^\infty (1+u)^{|S|-1}(1+u)^{M|S|/p}\exp(-uM|S|/p)\mathrm{d}u$$

$$= \exp(M|S|/p) \int_1^\infty u^{M|S|/p+|S|-1}\exp(-uM|S|/p)\mathrm{d}u$$

$$\leq \exp(M|S|/p) \int_0^\infty u^{M|S|/p+|S|-1}\exp(-uM|S|/p)\mathrm{d}u$$

$$= \frac{\exp(M|S|/p)}{(M|S|/p)^{M|S|/p+|S|}}\Gamma(M|S|/p + |S|)$$

$$= \frac{\exp(M|S|/p)}{(M|S|/p)^{M|S|/p+|S|}} (M|S|/p+1)^{|S|-1} \Gamma(M|S|/p+1)(1+o(1))$$

$$= \frac{\exp(M|S|/p)}{(M|S|/p)^{M|S|/p+|S|}} (M|S|/p+1)^{|S|-1} \sqrt{2\pi M|S|/p} \Big(\frac{M|S|/p}{e}\Big)^{M|S|/p} (1+o(1))$$

$$= \sqrt{2\pi}(M|S|/p)^{-1/2} \Big(1+\frac{p}{M|S|}\Big)^{|S|-1} (1+o(1)),$$

where the last three steps are from Stirling's approximation using $M \to \infty$.

By $M \to \infty$ and Assumption A.1, we have

$$\lim_{N_0 \to \infty} \Big(\frac{N_0}{M}\Big)^{p-|S|} \int_1^\infty u^{|S|-1} \mathrm{E}\Big[\mathbb{1}\Big(V_k > \frac{M}{N_0}u\Big)\Big([(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\ldots,Z_p)\to z} - \phi\|_\infty^d]\wedge 1\Big)^{p-|S|}\Big] \mathrm{d}u$$

$$= 0. \tag{B.13}$$

Combining (B.10), (B.11), (B.12), (B.13) by (B.9) yields

$$\limsup_{N_0\to\infty} \Big(\frac{N_0}{M}\Big)^p \mathrm{P}\Big(W_{\max} < V_{\max} \leq \eta_N \frac{M}{N_0}, \max_{k\notin S}\|\phi(x)-\phi(Z_k)\| < 4\|\widehat{\phi}-\phi\|_\infty/\epsilon' \leq \delta\Big)$$

$$\leq \frac{1}{p}|S|\Big[\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\Big]^{|S|}\mathbb{1}(|S|=p) = \Big[\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\Big]^p\mathbb{1}(|S|=p).$$

For the second term in (B.8), by the Chernoff bound,

$$\limsup_{N_0\to\infty} \Big(\frac{N_0}{M}\Big)^p \mathrm{P}\Big(V_{\max} > \eta_N \frac{M}{N_0}\Big) \leq |S|\limsup_{N_0\to\infty} \Big(\frac{N_0}{M}\Big)^p \mathrm{P}\Big(U_{(M)} > \eta_N \frac{M}{N_0}\Big) = 0.$$

For the third term in (B.8),

$$\mathrm{P}\Big(4\|\widehat{\phi}-\phi\|_\infty/\epsilon' > \delta\Big) \leq \delta^{-pd}(4/\epsilon')^{pd}\mathrm{E}[\|\widehat{\phi}-\phi\|_\infty^{pd}]$$

$$\leq \delta^{-pd}(4/\epsilon')^{pd}\mathrm{E}[\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\ldots,Z_p)\to z}-\phi\|_\infty^{pd}].$$

By Assumption A.1,

$$\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p\mathrm{P}\Big(4\|\widehat{\phi}-\phi\|_\infty/\epsilon' > \delta\Big) = 0.$$

Then we obtain for any $S \subset [\![p]\!]$,

$$\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p\mathrm{P}\Big(\nu_0(B_{\phi(Z_k),(1-\epsilon')\|\phi(x)-\phi(Z_k)\|}) \leq \nu_0(B_{\phi(Z_k),\Phi_M(Z_k)}) \text{ for } k \in S,$$

$$4\|\widehat{\phi}-\phi\|_\infty > \epsilon'\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|\Big)$$

$$\leq \Big[\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\Big]^p\mathbb{1}(|S|=p),$$

and then by (B.7),

$$\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p\mathrm{P}\Big(Z_1,\ldots,Z_p \in A_\phi(x)\Big) \leq \Big[\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\Big]^p.$$

By ϵ,ϵ' are arbitrary, we obtain

$$\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p\mathrm{P}\Big(Z_1,\ldots,Z_p \in A_\phi(x)\Big) \leq \Big(\frac{f_1(\phi(x))}{f_0(\phi(x))}\Big)^p.$$

A matched lower bound is directly from the Hölder inequality.

Then we obtain
$$\lim_{N_0 \to \infty} \left(\frac{N_0}{M}\right)^p P\left(Z_1, \ldots, Z_p \in A_\phi(x)\right) = \left(\frac{f_1(\phi(x))}{f_0(\phi(x))}\right)^p.$$

Part III. We consider the simple case where $p = 1$ and Assumption A.2 holds. We only consider the case where $f_1(\phi(x)) > 0$, while the case where $f_1(\phi(x)) = 0$ is similar.

Upper bound. Note that

$$P\left(Z_1 \in A_\phi(x)\right) = P\left(\|\widehat{\phi}(x) - \widehat{\phi}(Z_1)\| \le \widehat{\Phi}_M(Z_1)\right)$$
$$\le P\left(\|\phi(x) - \phi(Z_1)\| - 2 \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \Phi_M(Z_1)\right).$$

The inequality is from the fact that under the event $\{\|\widehat{\phi}(x) - \widehat{\phi}(Z_1)\| \le \widehat{\Phi}_M(Z_1)\}$, if $\|\phi(x) - \phi(Z_1)\| - 2 \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi}-\phi)(s) - (\widehat{\phi}-\phi)(t)\| > \Phi_M(Z_1)$, then there exists a set $S \subset \llbracket N_0 \rrbracket$ such that $|S| \ge M$ and for any $i \in S$, $\|\phi(x)-\phi(Z_1)\| - 2 \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi}-\phi)(s) - (\widehat{\phi} - \phi)(t)\| > \|\phi(X_i) - \phi(Z_1)\|$. For these $i \in S$, we then have $\|\widehat{\phi}(x) - \widehat{\phi}(Z_1)\| > \|\widehat{\phi}(X_i) - \widehat{\phi}(Z_1)\|$ since $\|\phi(x) - \phi(Z_1)\| > \|\phi(X_i) - \phi(Z_1)\|$. Then $\|\widehat{\phi}(x) - \widehat{\phi}(Z_1)\| > \widehat{\Phi}_M(Z_1)$ using $|S| \ge M$, which contradicts the event we assume.

For any $\epsilon \in (0, 1)$, we decompose as

$$P\left(\|\phi(x) - \phi(Z_1)\| - 2 \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \Phi_M(Z_1)\right)$$
$$= P\left(\|\phi(x) - \phi(Z_1)\| - 2 \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \Phi_M(Z_1),\right.$$
$$\left. 2 \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \epsilon\|\phi(x) - \phi(Z_1)\|\right)$$
$$+ P\left(\|\phi(x) - \phi(Z_1)\| - 2 \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \Phi_M(Z_1),\right.$$
$$\left. 2 \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| > \epsilon\|\phi(x) - \phi(Z_1)\|\right). \tag{B.14}$$

For the first term in (B.14),

$$P\left(\|\phi(x) - \phi(Z_1)\| - 2 \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \Phi_M(Z_1),\right.$$
$$\left. 2 \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \epsilon\|\phi(x) - \phi(Z_1)\|\right)$$
$$\le P\left((1 - \epsilon)\|\phi(x) - \phi(Z_1)\| \le \Phi_M(Z_1)\right),$$

and then can be handled in the same way as the upper bound part in Part I.

For the second term in (B.14), note that for any $u \in (0, 1)$, conditional on $\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) = u$ and under $\|\phi(x) - \phi(Z_1)\| \le \delta$, we have $u \le (1 + \epsilon)f_1(\phi(x))V_m\|\phi(x) - \phi(Z_1)\|^m$, and then $\|\phi(x) - \phi(Z_1)\| \ge [u/((1 + \epsilon)f_1(\phi(x))V_m)]^{1/m}$. Then for any $u \in (0, 1)$,

$$P\left(\sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| > \epsilon\|\phi(x) - \phi(Z_1)\|, \|\phi(x) - \phi(Z_1)\| \le \delta\right.$$

41

$$\left| \nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) = u \right)$$

$$\leq \mathrm{P}\Big(\sup_{\|\phi(s)-\phi(t)\|\leq\|\phi(x)-\phi(Z_1)\|} \sup_{z\in\mathcal{Z}} \|(\widehat{\phi}_{Z_1\to z} - \phi)(s) - (\widehat{\phi}_{Z_1\to z} - \phi)(t)\| > \epsilon\|\phi(x)-\phi(Z_1)\|, \|\phi(x)-\phi(Z_1)\| \leq \delta$$

$$\left| \nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) = u \right)$$

$$\leq T_\epsilon([u/((1+\epsilon)f_1(\phi(x))V_m)]^{1/m}),$$

where the last step is from the fact that $\sup_{z\in\mathcal{Z}}\|(\widehat{\phi}_{Z_1\to z} - \phi)(s) - (\widehat{\phi}_{Z_1\to z} - \phi)(t)\|$ does not depend on Z_1 together with Assumption A.2.

Then

$$\mathrm{P}\Big(\|\phi(x)-\phi(Z_1)\| - 2 \sup_{\|\phi(s)-\phi(t)\|\leq\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi}-\phi)(s) - (\widehat{\phi}-\phi)(t)\| \leq \Phi_M(Z_1),$$

$$2 \sup_{\|\phi(s)-\phi(t)\|\leq\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi}-\phi)(s) - (\widehat{\phi}-\phi)(t)\| > \epsilon\|\phi(x)-\phi(Z_1)\|\Big)$$

$$\leq \mathrm{P}\Big(2 \sup_{\|\phi(s)-\phi(t)\|\leq\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi}-\phi)(s) - (\widehat{\phi}-\phi)(t)\| > \epsilon\|\phi(x)-\phi(Z_1)\|\Big)$$

$$\leq \mathrm{P}\Big(2 \sup_{\|\phi(s)-\phi(t)\|\leq\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi}-\phi)(s) - (\widehat{\phi}-\phi)(t)\| > \epsilon\|\phi(x)-\phi(Z_1)\|, \|\phi(x)-\phi(Z_1)\| \leq \delta\Big)$$

$$+ \mathrm{P}\Big(2 \sup_{\|\phi(s)-\phi(t)\|\leq\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi}-\phi)(s) - (\widehat{\phi}-\phi)(t)\| > \epsilon\|\phi(x)-\phi(Z_1)\|, \|\phi(x)-\phi(Z_1)\| > \delta\Big)$$

$$\leq \int_0^1 T_{\epsilon/2}([u/((1+\epsilon)f_1(\phi(x))V_m)]^{1/m})\mathrm{d}u + \mathrm{P}\Big(\|\widehat{\phi}-\phi\|_\infty > \epsilon\delta/4 \Big)$$

$$= (1+\epsilon)f_1(\phi(x))V_m \int_0^{1/((1+\epsilon)f_1(\phi(x))V_m)} T_{\epsilon/2}(u^{1/m})\mathrm{d}u + \mathrm{P}\Big(\|\widehat{\phi}-\phi\|_\infty > \epsilon\delta/4 \Big).$$

By Assumption A.2 and ϵ is arbitrary, using (B.14), we obtain

$$\limsup_{N_0\to\infty} \frac{N_0}{M}\mathrm{P}\Big(Z_1 \in A_\phi(x) \Big) \leq \frac{f_1(\phi(x))}{f_0(\phi(x))}.$$

Lower bound. For any $\epsilon \in (0,1)$,

$$\mathrm{P}\Big(Z_1 \in A_\phi(x) \Big) = \mathrm{P}\Big(\|\widehat{\phi}(x) - \widehat{\phi}(Z_1)\| \leq \widehat{\Phi}_M(Z_1) \Big)$$

$$\geq \mathrm{P}\Big(\|\widehat{\phi}(x) - \widehat{\phi}(Z_1)\| \leq \widehat{\Phi}_M(Z_1), \sup_{\delta\geq\|\phi(x)-\phi(Z_1)\|} \delta^{-1} \sup_{\|\phi(s)-\phi(t)\|\leq\delta} \|(\widehat{\phi}-\phi)(s) - (\widehat{\phi}-\phi)(t)\| \leq \epsilon \Big)$$

$$\geq \mathrm{P}\Big((1+\epsilon)\|\phi(x)-\phi(Z_1)\| \leq (1-\epsilon)\Phi_M(Z_1), \sup_{\delta\geq\|\phi(x)-\phi(Z_1)\|} \delta^{-1} \sup_{\|\phi(s)-\phi(t)\|\leq\delta} \|(\widehat{\phi}-\phi)(s) - (\widehat{\phi}-\phi)(t)\| \leq \epsilon \Big).$$

The last inequality is from the fact that under the event $\{(1+\epsilon)\|\phi(x)-\phi(Z_1)\| \leq (1-\epsilon)\Phi_M(Z_1)\}$, there exists a set $S \subset [\![N_0]\!]$ such that $|S| \geq N_0 - M$ and for any $i \in S$, $(1+\epsilon)\|\phi(x)-\phi(Z_1)\| \leq (1-\epsilon)\|\phi(X_i)-\phi(Z_1)\|$. Under the event that $\{\sup_{\delta\geq\|\phi(x)-\phi(Z_1)\|}\delta^{-1}\sup_{\|\phi(s)-\phi(t)\|\leq\delta}\|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\| \leq \epsilon\}$, for these $i \in S$, we then have $\|\widehat{\phi}(x)-\widehat{\phi}(Z_1)\| \leq \|\widehat{\phi}(X_i)-\widehat{\phi}(Z_1)\|$ since $\|\phi(x)-\phi(Z_1)\| \leq \|\phi(X_i)-\phi(Z_1)\|$ for $i \in S$. Then $\|\widehat{\phi}(x)-\widehat{\phi}(Z_1)\| \leq \widehat{\Phi}_M(Z_1)$.

Then

$$\mathrm{P}\Big(Z_1 \in A_\phi(x) \Big) \geq \mathrm{P}\Big((1+\epsilon)\|\phi(x)-\phi(Z_1)\| \leq (1-\epsilon)\Phi_M(Z_1) \Big)$$

$$-\mathrm{P}\Big(\sup_{\delta\geq\|\phi(x)-\phi(Z_1)\|}\delta^{-1}\sup_{\|\phi(s)-\phi(t)\|\leq\delta}\|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\|>\epsilon\Big).$$

The first term can be handled in the same way as the lower bound part in Part I. The second term can be handled in the same way as the second term in (B.14) in this proof. Then we can obtain a matched lower bound.

Part IV. We then consider the general case where p is a fixed positive integer and Assumption A.2 holds. We only consider the case where $f_1(\phi(x))>0$, while the case where $f_1(\phi(x))=0$ is similar.

For any $\epsilon'\in(0,1)$, we have

$$\mathrm{P}\Big(Z_1,\ldots,Z_p\in A_\phi(x)\Big)=\mathrm{P}\Big(\|\widehat{\phi}(x)-\widehat{\phi}(Z_k)\|\leq\widehat{\Phi}_M(Z_k),\forall k\in[\![p]\!]\Big)$$

$$\leq\mathrm{P}\Big(\|\phi(x)-\phi(Z_k)\|-2\sup_{\|\phi(s)-\phi(t)\|\leq\|\phi(x)-\phi(Z_k)\|}\|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\|\leq\Phi_M(Z_k),\forall k\in[\![p]\!]\Big)$$

$$=\sum_{S\subset[\![p]\!]}\mathrm{P}\Big(\|\phi(x)-\phi(Z_k)\|-2\sup_{\|\phi(s)-\phi(t)\|\leq\|\phi(x)-\phi(Z_k)\|}\|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\|\leq\Phi_M(Z_k),$$

$$2\sup_{\|\phi(s)-\phi(t)\|\leq\|\phi(x)-\phi(Z_k)\|}\|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\|\leq\epsilon'\|\phi(x)-\phi(Z_k)\|\text{ for }k\in S,$$

$$2\sup_{\|\phi(s)-\phi(t)\|\leq\|\phi(x)-\phi(Z_k)\|}\|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\|>\epsilon'\|\phi(x)-\phi(Z_k)\|\text{ for }k\notin S\Big)$$

$$\leq\sum_{S\subset[\![p]\!]}\mathrm{P}\Big((1-\epsilon')\|\phi(x)-\phi(Z_k)\|\leq\Phi_M(Z_k)\text{ for }k\in S,$$

$$2\|\phi(x)-\phi(Z_k)\|^{-1}\sup_{\|\phi(s)-\phi(t)\|\leq\|\phi(x)-\phi(Z_k)\|}\|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\|>\epsilon'\text{ for }k\notin S\Big).$$

If $|S|=p$, we have in the same way as the upper bound part in Part III that for any $\epsilon\in(0,1)$,

$$\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p\mathrm{P}\Big((1-\epsilon')\|\phi(x)-\phi(Z_k)\|\leq\Phi_M(Z_k)\text{ for }k\in S,$$

$$2\|\phi(x)-\phi(Z_k)\|^{-1}\sup_{\|\phi(s)-\phi(t)\|\leq\|\phi(x)-\phi(Z_k)\|}\|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\|>\epsilon'\text{ for }k\notin S\Big)$$

$$\leq\Big[\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\Big]^p.$$

Now we consider $|S|<p$. Recall that $W_k=\nu_0(B_{\phi(Z_k),(1-\epsilon')\|\phi(x)-\phi(Z_k)\|})$ and $V_k=\nu_0(B_{\phi(Z_k),\Phi_M(Z_k)})$ for any $k\in[\![p]\!]$. Fix $S\subset[\![p]\!]$. Recall that $W_{\max}=\max_{k\in S}W_k$ and $V_{\max}=\max_{k\in S}V_k$. We have

$$\mathrm{P}\Big((1-\epsilon')\|\phi(x)-\phi(Z_k)\|\leq\Phi_M(Z_k)\text{ for }k\in S,$$

$$2\|\phi(x)-\phi(Z_k)\|^{-1}\sup_{\|\phi(s)-\phi(t)\|\leq\|\phi(x)-\phi(Z_k)\|}\|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\|>\epsilon'\text{ for }k\notin S\Big)$$

$$\leq\mathrm{P}\Big(W_{\max}<V_{\max},2\min_{k\notin S}\|\phi(x)-\phi(Z_k)\|^{-1}\sup_{\|\phi(s)-\phi(t)\|\leq\|\phi(x)-\phi(Z_k)\|}\|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\|>\epsilon'\Big)$$

$$\leq\mathrm{P}\Big(W_{\max}<V_{\max}\leq\eta_N\frac{M}{N_0},2\min_{k\notin S}\|\phi(x)-\phi(Z_k)\|^{-1}\sup_{\|\phi(s)-\phi(t)\|\leq\|\phi(x)-\phi(Z_k)\|}\|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\|>\epsilon'\Big)$$

43

$$+ \mathrm{P}\Big(V_{\max} > \eta_N \frac{M}{N_0}\Big).$$

Note that
$$2\min_{k\notin S}\|\phi(x) - \phi(Z_k)\|^{-1} \sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_k)\|} \|(\widehat\phi - \phi)(s) - (\widehat\phi - \phi)(t)\| > \epsilon'$$

implies that
$$(\max_{k\notin S}\|\phi(x) - \phi(Z_k)\|)^{-1} \sup_{\|\phi(s)-\phi(t)\|\le\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|} \|(\widehat\phi - \phi)(s) - (\widehat\phi - \phi)(t)\| > \epsilon'/2.$$

Then
$$\mathrm{P}\Big(W_{\max} < V_{\max} \le \eta_N \frac{M}{N_0}, 2\min_{k\notin S}\|\phi(x) - \phi(Z_k)\|^{-1} \sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_k)\|} \|(\widehat\phi - \phi)(s) - (\widehat\phi - \phi)(t)\| > \epsilon'\Big)$$

$$\le \mathrm{P}\Big(\Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)\frac{f_0(\phi(x))}{f_1(\phi(x))} \max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}) < V_{\max},$$

$$(\max_{k\notin S}\|\phi(x) - \phi(Z_k)\|)^{-1} \sup_{\|\phi(s)-\phi(t)\|\le\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|} \|(\widehat\phi - \phi)(s) - (\widehat\phi - \phi)(t)\| > \epsilon'/2\Big).$$

In the same way as the upper bound part in Part II, it suffices to show that
$$\lim_{N_0\to\infty} \Big(\frac{N_0}{M}\Big)^{p-|S|}\mathrm{P}\Big((\max_{k\notin S}\|\phi(x) - \phi(Z_k)\|)^{-1} \sup_{\|\phi(s)-\phi(t)\|\le\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|} \|(\widehat\phi - \phi)(s) - (\widehat\phi - \phi)(t)\| > \epsilon'/2\Big) = 0.$$

For any $u \in (0,1)$, conditional on $\max_{k\notin S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}) = u$ and under $\max_{k\notin S}\|\phi(x) - \phi(Z_k)\| \le \delta$, we have
$$\max_{k\notin S}\|\phi(x) - \phi(Z_k)\| \ge [u/((1+\epsilon)f_1(\phi(x))V_m)]^{1/m}.$$

Then by Assumption A.2,
$$\Big(\frac{N_0}{M}\Big)^{p-|S|}\mathrm{P}\Big((\max_{k\notin S}\|\phi(x) - \phi(Z_k)\|)^{-1} \sup_{\|\phi(s)-\phi(t)\|\le\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|} \|(\widehat\phi - \phi)(s) - (\widehat\phi - \phi)(t)\| > \epsilon'/2\Big)$$

$$= \Big(\frac{N_0}{M}\Big)^{p-|S|} \int_0^1 (p-|S|)u^{p-|S|-1}\mathrm{P}\Big((\max_{k\notin S}\|\phi(x) - \phi(Z_k)\|)^{-1} \sup_{\|\phi(s)-\phi(t)\|\le\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|}$$

$$\|(\widehat\phi - \phi)(s) - (\widehat\phi - \phi)(t)\| > \epsilon'/2, \max_{k\notin S}\|\phi(x) - \phi(Z_k)\| \le \delta \,\Big|\, \max_{k\notin S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}) = u\Big)\mathrm{d}u$$

$$+ \Big(\frac{N_0}{M}\Big)^{p-|S|}\mathrm{P}\Big(\|\widehat\phi - \phi\|_\infty > \epsilon'\delta/4\Big)$$

$$\le \Big(\frac{N_0}{M}\Big)^{p-|S|}\Big[\int_0^1 (p-|S|)u^{p-|S|-1}T_{\epsilon'}([u/(\|f_1\|_\infty V_m)]^{1/m})\mathrm{d}u + \mathrm{P}\Big(\|\widehat\phi - \phi\|_\infty > \epsilon'\delta/4\Big)\Big] = o(1).$$

Then
$$\limsup_{N_0\to\infty} \Big(\frac{N_0}{M}\Big)^p\mathrm{P}\Big(Z_1,\ldots,Z_p \in A_\phi(x)\Big) \le \Big[\Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\Big]^p.$$

By ϵ, ϵ' are arbitrary, we obtain
$$\limsup_{N_0\to\infty} \Big(\frac{N_0}{M}\Big)^p\mathrm{P}\Big(Z_1,\ldots,Z_p \in A_\phi(x)\Big) \le \Big(\frac{f_1(\phi(x))}{f_0(\phi(x))}\Big)^p.$$

A matched lower bound is directly from the Hölder inequality.

B.4 Proof of Theorem A.2

Consider any $\epsilon \in (0,1)$ be given. From Assumption A.3, $\mathcal{X}$ is compact, and then $\mathcal{Z}$ is also compact. Since f_0, f_1 are continuous over their compact supports, they are uniformly continuous, that is, there exists $\delta > 0$ such that for any $x, z \in \mathcal{Z}$ with $\|\phi(z) - \phi(x)\| \le 3\delta$, we have $|f_1(\phi(z)) - f_1(\phi(x))| \le \epsilon^2$, and for any $x, z \in \mathcal{X}$ with $\|\phi(z) - \phi(x)\| \le 3\delta$, we have $|f_0(\phi(z)) - f_0(\phi(x))| \le \epsilon^2$.

Let $\mathcal{E}_1 = \{x : f_1(\phi(x)) \le \epsilon\}$, $\mathcal{E}_2 = \{x : f_0(\phi(x)) \le \epsilon \text{ or } \mathrm{dist}(x, \partial\mathcal{Z}) \vee \mathrm{dist}(x, \partial\mathcal{X}) \le 3\delta\}$. We then seperate the proof into three cases. In the following, it suffices to consider x such that $f_0(\phi(x)) > 0$ since we are considering L_p risk.

Case I. $x \notin \mathcal{E}_1 \cup \mathcal{E}_2$. In this case we have $f_0(\phi(x)), f_1(\phi(x)) > \epsilon$. Then for any $z \in \mathcal{X}$ with $\|\phi(z) - \phi(x)\| \le 3\delta$, we have $z \in \mathcal{Z}$ by the definition of $\mathcal{E}_2$ and then $|f_0(\phi(z)) - f_0(\phi(x))| \le \epsilon f_0(\phi(x))$ and $|f_1(\phi(z)) - f_1(\phi(x))| \le \epsilon f_1(\phi(x))$.

Proceeding as in the proof of Case I in Lemma A.3, we obtain

$$\lim_{N_0 \to \infty} \sup_{x \notin \mathcal{E}_1 \cup \mathcal{E}_2} \mathrm{E}\left[\left|\widehat{r}_\phi(x) - r(\phi(x))\right|^p\right] = 0,$$

and then

$$\lim_{N_0 \to \infty} \mathrm{E}\left[\left|\widehat{r}_\phi(X) - r(\phi(X))\right|^p \mathbb{1}\left(X \notin \mathcal{E}_1 \cup \mathcal{E}_2\right)\right] = \lim_{N_0 \to \infty} \mathrm{E}\left[\mathrm{E}\left[\left|\widehat{r}_\phi(X) - r(\phi(X))\right|^p\right] \mathbb{1}\left(X \notin \mathcal{E}_1 \cup \mathcal{E}_2\right)\right]$$

$$\le \lim_{N_0 \to \infty} \mathrm{E}\left[\sup_{x \notin \mathcal{E}_1 \cup \mathcal{E}_2} \mathrm{E}\left[\left|\widehat{r}_\phi(x) - r(\phi(x))\right|^p\right] \mathbb{1}\left(X \notin \mathcal{E}_1 \cup \mathcal{E}_2\right)\right] = 0.$$

Case II. $x \in \mathcal{E}_1 \setminus \mathcal{E}_2$. In this case we have $f_0(\phi(x)) > \epsilon$. Then for any $z \in \mathcal{X}$ with $\|\phi(z) - \phi(x)\| \le 3\delta$, we have $z \in \mathcal{Z}$ by the definition of $\mathcal{E}_2$ and then $|f_0(\phi(z)) - f_0(\phi(x))| \le \epsilon f_0(\phi(x))$ and $f_1(\phi(z)) \le \epsilon + \epsilon^2$.

Proceeding as in the proof of Case II in Lemma A.3, we obtain

$$\lim_{N_0 \to \infty} \sup_{x \in \mathcal{E}_1 \setminus \mathcal{E}_2} \mathrm{E}\left[\left|\widehat{r}_\phi(x) - r(\phi(x))\right|^p\right] = 0,$$

and then

$$\lim_{N_0 \to \infty} \mathrm{E}\left[\left|\widehat{r}_\phi(X) - r(\phi(X))\right|^p \mathbb{1}\left(X \notin \mathcal{E}_1 \setminus \mathcal{E}_2\right)\right] = 0.$$

Case III. $x \in \mathcal{E}_2$. In this case we have $f_0(\phi(x)) \le \epsilon$ or $\mathrm{dist}(x, \partial\mathcal{Z}) \vee \mathrm{dist}(x, \partial\mathcal{X}) \le 3\delta$. Since $\mathcal{X}$ is compact, the surface areas of $\mathcal{X}$ and $\mathcal{Z}$ are bounded, and r is bounded uniformly, we have

$$\limsup_{N_0 \to \infty} \mathrm{E}\left[\left|\widehat{r}_\phi(X) - r(\phi(X))\right|^p \mathbb{1}\left(X \in \mathcal{E}_2\right)\right] \lesssim \mathrm{P}\left(X \in \mathcal{E}_2\right) \lesssim \epsilon + \delta.$$

Since ϵ is arbitrary and δ can be taken arbitrary small, we obtain

$$\lim_{N_0 \to \infty} \mathrm{E}\left[\left|\widehat{r}_\phi(X) - r(\phi(X))\right|^p \mathbb{1}\left(X \in \mathcal{E}_2\right)\right] = 0.$$

Combining the above three cases completes the proof.

B.5 Proof of Lemma A.4

Note that for any $x \in \mathcal{X}$,

$$\frac{f_{\phi,X \mid D=1}(\phi(x))}{f_{\phi,X \mid D=0}(\phi(x))} = \frac{\mathrm{P}(D=1 \mid \phi(X)=\phi(x))\mathrm{P}(D=0)}{\mathrm{P}(D=0 \mid \phi(X)=\phi(x))\mathrm{P}(D=1)},$$

and

$$\frac{f_{X \mid D=1}(x)}{f_{X \mid D=0}(x)} = \frac{\mathrm{P}(D=1 \mid X=x)\mathrm{P}(D=0)}{\mathrm{P}(D=0 \mid X=x)\mathrm{P}(D=1)}.$$

This completes the proof.

References

Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74(1):235–267.

Abadie, A. and Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business and Economic Statistics*, 29(1):1–11.

Abadie, A. and Imbens, G. W. (2012). A martingale representation for matching estimators. *Journal of the American Statistical Association*, 107(498):833–843.

Belloni, A., Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics*, 186(2):345–366.

Bickel, P. J. (2004). Rietz lecture: The frontiers of statistics and computer science. Personal communication.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2015). Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association*, 110(512):1753–1769.

Cattaneo, M. D., Crump, R. K., Farrell, M. H., and Feng, Y. (2023). On binscatter. *arXiv preprint arXiv:1902.09608*.

Cattaneo, M. D. and Farrell, M. H. (2013). Optimal convergence rates, bahadur representation, and asymptotic normality of partitioning estimators. *Journal of Econometrics*, 174(2):127–143.

Cattaneo, M. D., Farrell, M. H., and Feng, Y. (2020). Large sample properties of partitioning-based series estimators. *Annals of Statistics*, 48(3):1718–1741.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, 66(2):315–331.

Huang, J. (2003). Local asymptotics for polynomial spline regression. *Annals of Statistics*, 31(5):1600–1635.

Joe, H. (2014). *Dependence Modeling with Copulas*. CRC press.

Li, Q. and Racine, J. S. (2023). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Lin, Z., Ding, P., and Han, F. (2023). Estimation based on nearest neighbor matching: from density ratio to average treatment effect. *Econometrica*, 91(6):2187–2217.

Lin, Z. and Han, F. (2022). On regression-adjusted imputation estimators of the average treatment effect. *arXiv preprint arXiv:2212.05424*.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79(1):147–168.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866.

Rosenbaum, P. R. (2005). An exact distribution-free test comparing two multivariate distributions based on adjacency. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(4):515–530.

Rosenbaum, P. R. (2020). *Design of Observational Studies*. Springer, 2nd edition.

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999). Adjusting for nonignorable dropout using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448):1096–1120.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1):1–21.

Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12:389–434.